Marcel Pelgrom

Analog-to-Digital Conversion

Third Edition

Springer

Analog-to-Digital Conversion

Marcel Pelgrom

Analog-to-Digital Conversion

Third Edition

Springer

Marcel Pelgrom
HELMOND
Noord-Brabant
The Netherlands

ISBN 978-3-319-44970-8 ISBN 978-3-319-44971-5 (eBook)
DOI 10.1007/978-3-319-44971-5

Library of Congress Control Number: 2016952089

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is Gewerbestrasse 11, 6330 Cham, Switzerland

Preface to the Third Edition

In this edition, the time-interleaved conversion is extended to a full chapter. The other chapters have been updated with the latest developments, and that has led to a considerable increase in pages for the Nyquist and sigma-delta conversion chapters. With these extensions, the third edition of the analog-to-digital conversion book became too voluminous. So the supporting components, technology, and systems chapters have been removed.

Stiphout, The Netherlands — Marcel Pelgrom
July 2016

Preface to the Second Edition

In the second edition, a number of errors have been corrected and a few topics have been updated. The most important change is the addition of many examples and exercises to assist students in understanding the material.

Stiphout, The Netherlands — Marcel Pelgrom
Summer 2012

Preface to the First Edition

A book is like a window that allows you to look into the world. The window is shaped by the author, and that makes every book present a unique view of the world. This is certainly true for this book. My views were shaped by the topics and the projects throughout my career. Even more so, this book reflects my own style of working and thinking.

In the chapters discussing the fundamental processes of conversion, sampling, and quantization, you will recognize my preoccupation with mathematics. I really enjoy finding an equation that properly describes the underlying mechanism. Nevertheless, mathematics is not a goal on its own: equations help to understand the way how variables are connected to the result. Real insight comes from understanding the physics and electronics. Electronics circuits are sometimes so ingenious that they almost feel like art. Yet, in the chapters on circuit design, I have tried to keep the descriptions pragmatic. The circuit diagrams are shown in their simplest form, but not simpler.... I do have private opinions on what works and what should not be applied. Most poor solutions have simply been left out; sometimes you might read a warning in the text on a certain aspect of an otherwise interesting circuit.

This book is based on lectures for graduate students who are novice in analog-to-digital design. In the classes, my aim is to bring the students to a level where they can read and interpret the advanced literature (such as *IEEE Journal of Solid-State Circuits*) and judge the reported results on their merits. Still that leaves a knowledge gap with the designer of analog-to-digital converters. For those experienced designers, this book may serve as a reference of principles and background.

Inevitably, this book has not only strong points but also weak points. There are still so many wonderful ideas that are not addressed here but certainly would deserve some space, but simply did not fit in this volume. Still I hope this book will let you experience the same thrill that all analog-to-digital designers feel, when they talk about their passion, because that is the goal of this book: to encourage you to proceed on the route toward even better analog-to-digital converters.

Stiphout, The Netherlands Marcel Pelgrom
Christmas 2009

Acknowledgments

Archimedes said: “Give me one fixed point and I will move the Earth.” Home has always served for me as the fixed point from which I could move forward in my work. I owe my wife Elisabeth a debt of gratitude for creating this wonderful home. She herself was once part of this semiconductor world and understands its crazy habits. Yet, the encouragement and support she gave me is invaluable.

This book reflects parts of my 33 years of work in the Philips Natuurkundig Laboratorium and its successor. If there is anything I would call “luck” in my life, it was the opportunity to work in this place. The creativity, energy, opportunities, and people in this laboratory are unique. It is not trivial to create such research freedom in a financially driven industry. My 7 years as a mixed-signal department head have taught me that. Therefore, I am truly grateful to those who served in the management of Philips Research and backed me when working on things outside the project scope or looking in unusual directions, namely, Theo van Kessel, Kees Wouda, Gerard Beenker, Hans Rijns, and Leo Warmerdam.

A laboratory is just as good as the people that work in it. In my career, I met a lot of extraordinary people. They formed and shaped my way of thinking and analyzing problems. They challenged my ideas, took the time to listen to my reasoning, and pointed me in promising directions. I am grateful for being able to use the insights and results of the mixed-signal circuits and systems group. Without the useful discussions and critical comments of the members of this group, this book would not exist. However, there are many more colleagues that have contributed in some form. Without the illusion of being complete, I want to express my gratitude for a pleasant collaboration with: Carel Dijkmans, Rudy van de Plassche, Eduard Stikvoort, Rob van der Grift, Arthur van Roermund, Erik van der Zwan, Peter Nuijten, Ed van Tuijl, Maarten Vertregt, Pieter Vorenkamp, Johan Verdaasdonk, Anton Welbers, Aad Duinmaijer, Jeannet van Rens, Klaas Bult, Govert Geelen, Stephane Barbu, Laurent Giry, Robert Meyer, Othmar Pfarkircher, Ray Speer, John Jennings, Bill Redman-White, Joost Briaire, Pieter van Beek, Raf Roovers, Lucien Breems, Robert van Veldhoven, Kathleen Philips, Bram Nauta, Hendrik van der Ploeg, Kostas Doris, Erwin Janssen, Robert Rutten, Violeta Petrescu, Harry Veendrick, Hans Tuinhout, Jan van der Linde, Peter van Leeuwen, and many others.

This book is based on the lectures in the Philips Center for Technical Training, at universities, and in the MEAD/EPFL courses. I want to thank Prof. Bram Nauta and Prof. Kofi Makinwa for giving me the opportunity to teach at the universities of Twente and Delft, prof. Bruce Wooley and Prof. Boris Murmann of Stanford University for their collaboration, and Prof. Gabor Temes and Dr. Vlado Valence for inviting me to lecture in the MEAD and EPFL courses.

A special word of thanks goes to all the students for their questions, remarks, and stimulating discussions.

Contents

List of Symbols

Symbol	Description	Unit symbol
A	Area	cm^2
C	Capacitance	F
C_{ox}	Oxide capacitance	F/cm^2
D_n	Diffusion coefficient of electrons	cm^2/sec
d_{ox}	Oxide thickness	Å=10^{-10}m
E	Electric field	V/cm
E_{Fn}	Fermi energy level of electrons	eV
E_{Fp}	Fermi energy level of holes	eV
E_G	Band gap energy	1.205 eV
E_i	Energy level of an intrinsic semiconductor	eV
f	Frequency	Hz
f_c	Clock frequency	Hz
f_{in}	Frequency of input signal	Hz
f_s	Sample rate	Hz
H	Transfer function	1
I	Large-signal current or DC	A
i	Small-signal current	A
J	Current density	A/cm^2
J_n	Electron current density	A/cm^2
J_p	Hole current density	A/cm^2
K	Substrate voltage influence on the threshold voltage	$\sqrt{V}$
k	Boltzmann's constant	1.38×10^{-23} J/°K
L	Length of transistor gate	µm
L_w	Inductance of a wire	H
N	Resolution	1
N_{int}	Multiplex factor/time interleave factor	1
N_a	Substrate doping concentration	cm^{-3}
n	Volume density of electrons	cm^{-3}

n_i	Intrinsic charge volume density (300 °K)	1.4×10^{10} cm^{-3}
p	Volume density of holes	cm^{-3}
p_{p0}	Volume density of holes in a p-substrate in equilibrium	cm^{-3}
R	Resistance	Ω
Q	Charge	C
q	Electron charge	1.6×10^{-19} C
T_x	Time period	s
T	Absolute temperature	°K
T_0	Reference temperature	°K
T_s	Sample period	s
t	Time as a running variable	s
V	Bias or DC potential	V
V_{DD}	Positive power supply	V
V_{DS}	Drain potential with respect to the source potential	V
V_{FB}	Flat-band voltage	V
V_G	Gate potential with respect to ground	V
V_{GS}	Gate potential with respect to the source potential	V
V_T	MOS transistor threshold voltage	V
v	Small-signal voltage	V
W	Width of transistor channel	μm
X, Y	General input and output variable	1
x, y, z	Dimension variables	μm
Z	Complex impedance	Ω
β	Current gain factor of MOS transistor: $W\beta_\square/L$	A/V^2
$\beta_\square$	Current gain factor of a square MOS transistor	A/V^2
ΔP	Difference between two parameters $P_1 - P_2$	
ϵ_0	Permittivity in vacuum	8.854×10^{-14} F/cm
$\epsilon_{ox}\epsilon_0$	Permittivity in silicon dioxide	3.45×10^{-13} F/cm
$\epsilon_s\epsilon_0$	Permittivity in silicon	10.5×10^{-13} F/cm
ϕ	Angular phase	degree
ϕ_F	Potential difference between intrinsic and hole Fermi level	V
μ	Mean value of a statistical distribution	
μ_0	Magnetic permeability in vacuum	$4\pi \times 10^{-7}$ H/m or N/A^{-2}
μ_n, μ_p	Mobility of electrons and holes	cm^2/Vs
π	Angular constant	3.14159
ψ	Electrostatic potential	V
ψ_B	Electrostatic potential at which strong inversion starts	V
ψ_s	Electrostatic potential at the interface	V
$\sigma_{\Delta P}$	Standard deviation of ΔP	
τ	Time constant	s
$\omega = 2\pi f$	Angular or radian frequency	rad/s

Reference Tables and Figures

Table 1 Multiplier abbreviations

Name	Abbreviation	Multiplier
Googol		10^{100}
Exa	E	10^{18}
Peta	P	10^{15}
Tera	T	10^{12}
Giga	G	10^{9}
Mega	M	10^{6}
kilo	k	10^{3}
hecto	h	10^{2}
deca	da	10
unity		1
deci	d	10^{-1}
centi	c	10^{-2}
milli	m	10^{-3}
micro	μ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}
femto	f	10^{-15}
atto	a	10^{-18}

Table 2 Probability that an experimental value exceeds the $z\sigma$ limit on one side in a normal distribution $P((x-\mu) > z\sigma)$ [6–8]

z	P	z	P	z	P	z	P	z	P	z	P
0	0.5000	1.0	0.1587	2.0	0.02275	3.0	$1350\ 10^{-6}$	4.0	$31671\ 10^{-9}$	5.0	$287\ 10^{-9}$
0.1	0.4602	1.1	0.1357	2.1	0.01786	3.1	$968\ 10^{-6}$	4.1	$20657\ 10^{-9}$	5.1	$170\ 10^{-9}$
0.2	0.4207	1.2	0.1151	2.2	0.01390	3.2	$687\ 10^{-6}$	4.2	$13346\ 10^{-9}$	5.2	$100\ 10^{-9}$
0.3	0.3821	1.3	0.0968	2.3	0.01072	3.3	$483\ 10^{-6}$	4.3	$8540\ 10^{-9}$	5.3	$57.9\ 10^{-9}$
0.4	0.3446	1.4	0.0808	2.4	0.00820	3.4	$337\ 10^{-6}$	4.4	$5413\ 10^{-9}$	5.4	$33.3\ 10^{-9}$
0.5	0.3085	1.5	0.0668	2.5	0.00621	3.5	$233\ 10^{-6}$	4.5	$3398\ 10^{-9}$	5.5	$19.0\ 10^{-9}$
0.6	0.2743	1.6	0.0548	2.6	0.00466	3.6	$159\ 10^{-6}$	4.6	$2112\ 10^{-9}$	5.6	$10.7\ 10^{-9}$
0.7	0.2420	1.7	0.0446	2.7	0.00347	3.7	$108\ 10^{-6}$	4.7	$1301\ 10^{-9}$	5.7	$5.99\ 10^{-9}$
0.8	0.2119	1.8	0.0359	2.8	0.00256	3.8	$72.3\ 10^{-6}$	4.8	$793\ 10^{-9}$	5.8	$3.32\ 10^{-9}$
0.9	0.1841	1.9	0.0287	2.9	0.00187	3.9	$48.1\ 10^{-6}$	4.9	$479\ 10^{-9}$	5.9	$1.82\ 10^{-9}$
1.0	0.1587	2.0	0.0228	3.0	0.00135	4.0	$31.7\ 10^{-6}$	5.0	$287\ 10^{-9}$	6.0	$0.987\ 10^{-9}$

The dual-sided rejection is easily found by doubling: $P(|x-\mu| > z\sigma) = 2 \times P((x-\mu) > z\sigma)$

Table 3 Basic MOS transistor equations

MOS transistor current	$I_D = \frac{\beta}{2}(V_{GS} - V_T + \lambda V_{DS})^2 = \frac{W\beta_\square}{2L}(V_{GS} - V_T + \lambda V_{DS})^2$
Current factor	$\beta = \frac{W}{L}\beta_\square = \frac{W}{L}\mu C_{ox}$
Threshold voltage	$V_T = V_{FB} + 2\phi_F + \frac{\sqrt{2\varepsilon_r\varepsilon_0 q N_a(2\phi_F + V_{SB})}}{C_{ox}}$ $= V_{FB} + 2\phi_F + \gamma\sqrt{2\phi_F + V_{SB}}$ $= V_T(V_{SB} = 0) + \gamma(\sqrt{2\phi_F + V_{SB}} - \sqrt{2\phi_F})$
Simple MOS current	$I_D = \frac{\beta}{2}(V_{GS} - V_T)^2$
Diode current	$I_{pn} = I_o(e^{\frac{qV_{pn}}{kT}} - 1)$
Transconductance	$g_m = \frac{dI_D}{dV_{GS}} = \frac{W\beta_\square}{L}(V_{GS} - V_T + \lambda V_{DS}) = \sqrt{\frac{2I_D W\beta_\square}{L}}$
Output conductance	$g_D = \frac{dI_D}{dV_{DS}} = \frac{\lambda W\beta_\square}{L}(V_{GS} - V_T + \lambda V_{DS}) = \lambda g_m$

Table 4 The values of NMOS and PMOS parameters are indicative for what is used in industry

Process	V_{DD}	d_{ox}	C_{ox}	$V_{T,n}$	$V_{T,p}$	$\beta_{\square,n}$	$\beta_{\square,p}$	$A_{VT,n}$	$A_{VT,p}$	$A_{\beta,np}$
Unit	Volt	nm	fF/μm^2	Volt	Volt	μA/V^2	μA/V^2	mVμm	mVμm	%μm
0.8 μm	3.3	15	2.3	0.6	−0.65	125	55	10.7	18.0	4
0.6 μm	3.3	13	2.7	0.65	−0.8	150	50	11.0	8.5	
0.5 μm	3.3	12	2.9	0.6	−0.6	130	36	9	10	1.8
0.35 μm	3.3	7.7	4.3	0.63	−0.6	190	46	8	7.4	2
0.25 μm	2.5	6	6.9	0.57	−0.53	235	53	6	6	1.5
0.18 μm	1.8	4	8.3	0.48	−0.5	300	80	6	5	1.6
0.13 μm	1.2	2.5	11	0.34	−0.36	590	135	5	5	1.6
90 nm (LP)	1.2	2.3	11.7	0.37	−0.39	550	160	4.5	3.5	
65 nm (LP)	1.2	2.2	12.6	0.32	−0.36	450	200	4.5	3.5	1.2
45 nm (LP)	1.1	1.7	16	0.39	−0.42	300	100	4	4	1.2
28 nm	1.0	1.0		0.35	−0.35			2	2	

Especially the 130–28 nm values depend on many technological effects and bias settings. The 28-nm data is based on process with a high-k factor dielectric with metal gate (*Source*: ITRS [13] and various publications)

About the Author

Marcel Pelgrom received his B.EE, M.Sc, and PhD from the University of Twente, Enschede, the Netherlands. In 1979, he joined the Philips Research Laboratories, where his research has covered topics such as charge-coupled devices, MOS matching properties, analog-to-digital conversion, digital image correlation, and various analog building block techniques. He has headed several project teams and was a team leader for high-speed analog-to-digital conversion. From 1996 until 2003, he was a department head for mixed-signal electronics. In addition to various activities concerning industry-academic relations, he is a Philips research fellow on the edge of design and technology. In 2003, he spent a sabbatical in Stanford University where he served as a consulting professor and in 2014 as a lecturer. From 2006 until 2013, he has been a member of the technical staff at NXP Semiconductors. Dr. Pelgrom was twice an IEEE distinguished lecturer and has written over 40 publications and seven book chapters, and he holds 35 US patents. He currently lectures at Delft University of Technology and the University of Twente and for MEAD Inc. He is the 2017 recipient of the IEEE field award Gustav R. Kirchhoff.

Chapter 1
Introduction

Analog-to-digital conversion is everywhere around us. In all forms of electronic equipment the analog-to-digital converter links our physical world to digital computing machines. This development has enabled the marvelous electronics functionality that has been introduced over the last 30 years, from mobile phone to internet, from medical imaging machines to hand-held television.

Analog signals are connected to physical quantities like length, weight, volume, or more appropriate for the electronics field: voltage, charge, current, and magnetic flux. Analog electronics circuits manipulate these quantities by means of operations such as amplification, attenuation, and modulation. Analog-only circuits can do a lot of signal processing in a cheap and well-established way. Many signal processing functions are so simple that analog processing serves the needs (audio amplification, filtering, radio, television, and first generation mobile phone). In more complex situations, analog processing, however, lacks required functionality and digital signal processing offers crucial improvements. The most important advantages of digital processing over analog processing are a perfect storage of digitized signals, unlimited signal-to-noise ratio, the option to carry out complex calculations, and the possibility to adapt the algorithm of the calculation to changing circumstances. If an application wants to use these advantages, analog signals have to be converted with high quality into a digital format in an early stage of the processing chain. And at the end of the digital processing the conversion has to be carried out in the reverse direction. The digital-to-analog translates the outcome of the signal processing into signals that can be rendered into a picture or sound. This makes analog-to-digital conversion a crucial element in the chain between the world of physical quantities and the rapidly increasing power of digital signal processing. Figure 1.1 shows the analog-to-digital converter (abbreviated A/D-converter or ADC) as the crucial element in a chain with combined analog and digital functionality.

The basic operations of the analog-to-digital conversion process are shown in Fig. 1.2. In the analog or physical domain, signals exist with continuous amplitudes and at every moment of time. Signals in the digital domain differ from analog

M. Pelgrom, *Analog-to-Digital Conversion*, DOI 10.1007/978-3-319-44971-5_1

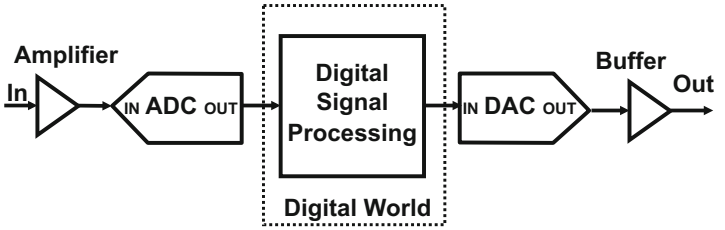


Fig. 1.1 The analog-to-digital and digital-to-analog converters are the ears and eyes of a digital system

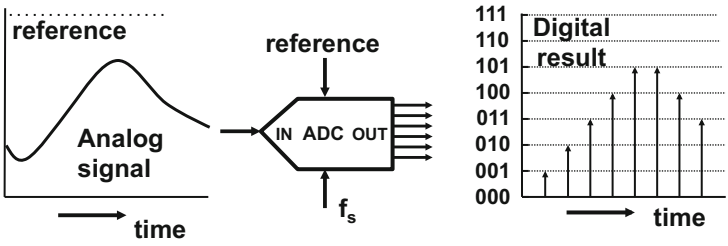


Fig. 1.2 Functions of the analog-to-digital converter: sampling, quantizing, and linking to a reference

signals, because digital signals are sampled and quantized. Sampled signals only have meaning at their sample moments as given by the sampling signal with repetition rate f_s.

The analog-to-digital converter compares the value of the analog signal at the sample moment to a fraction of the reference quantity (voltage, current, charge, time). **The accuracy of this analog fraction determines the accuracy of the conversion.** This analog fraction is approximated in the digital domain, where the digital code is a fraction of the available word width. The analog-to-digital converter tries to find the digital code that gives an optimum match between these ratios at every sample moment.

Digital signals are arithmetic quantities that are only meaningful in the physical world when a physical reference value is assigned to the digital number range. The reference value is therefore crucial for both the conversion from the analog domain into a digital number and the conversion from digital numbers towards physical quantities. Figure 1.3 shows a possible reconstruction result of a digital signal. Comparison of the analog input in Fig. 1.2 and the analog output in Fig. 1.3 shows the inevitable rounding error in every conversion process. A better match can be obtained by sampling at smaller time intervals (increasing the sample rate) or using denser spaced digital levels (increasing the resolution). The strive towards more sample rate and more resolution is a continuous driving force in analog-to-digital conversion research. Some topologies are more optimum for resolution, other

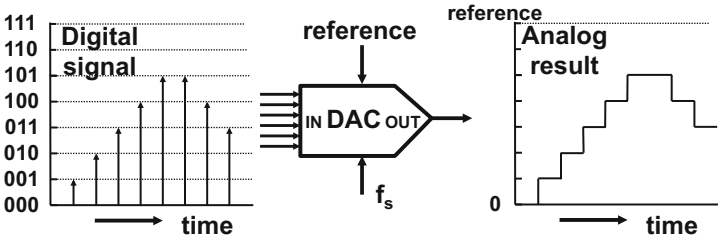


Fig. 1.3 Functions of the analog-to-digital converter: sampling, quantizing, and linking to a reference

Table 1.1 Key functions in analog-to-digital conversion

Analog-to-digital	Digital-to-analog
Time discretization or sampling	Holding the signal
Amplitude discretization or quantization	Amplitude restoration
Compare to a reference quantity	Derive from a reference quantity

topologies are more beneficial to increase the speed. But, increasing resolution or speed always requires more power.

In this paragraph three main functions of the analog-to-digital conversion and digital-to-analog conversion have been introduced, see Table 1.1. These functions will be visible in each stage of the discussion of analog-to-digital conversion and are reflected in the setup of this book.

1.1 About this Book

An analog-to-digital converter and a digital-to-analog converter are electronics circuits that are designed and fabricated in silicon IC technology. Several books have been published in the field of analog-to-digital conversion. One of the first books was published by Seitzer [1] in 1983, describing the basic principles. Van de Plassche [2] in 1994 and 2003 and Razavi [3] in 1994 discuss extensively bipolar and CMOS realizations. Jespers [4] and Maloberti [5] address the theme on a graduate level. These textbooks review the essential aspects and focus on general principles, circuit realizations, and their merits.

This present book reviews the fundamental principles and includes new insights coming from a different balance between technology advances and system requirements. The state of the art is described as far as it fits within the scope of this book. Statistical analysis is introduced. The main focus in this book is on CMOS realizations.

In Chaps. 2–6 the three basic functions for conversion are analyzed. Chapter 2 describes the sampling process and gives guidelines for the different choices around

sampling in analog-to-digital conversion design. The design challenges around the sampling process are discussed in the design of sample-and-hold circuits in Chap. 3. Both sampling and quantization are non-linear operations, resulting in the required conversion but accompanied by some undesired behavior. The combination of sampling and quantization comes with a fundamental error: the quantization error. The attainable performance of every analog-to-digital conversion is fundamentally limited by this error as is described in Chap. 4.

Every analog-to-digital converter and digital-to-analog converter is only as accurate as the (relative) accuracy of its components. Chapter 5 summarizes some of the component properties where they are relevant to analog-to-digital conversion. Stable reference values are a prerequisite for a quality conversion process. Chapter 6 deals with the generation and handling of reference voltages.

The main question addressed in this book is how to construct converter circuits. Chapter 7 describes the basics of digital-to-analog converter circuit design and some implementations of digital-to-analog converters.

Earlier the fundamental quantization error was mentioned. Two directions exist to improve speed and resolution: increasing the number of levels for comparison and oversampling. The design of the first type of "Nyquist-rate" analog-to-digital converters is elaborated in Chap. 8. In Chap. 9 the advantages and disadvantages of boosting the speed of these converters by means of time-interleaving analog-to-digital conversion topologies are discussed.

Oversampling and sigma-delta conversion improve the resolution by taking sampling at a speed much higher than the required bandwidth and improving the resolution from the combination of many samples. These converters are discussed in Chap. 10.

The measurement methods for analog-to-digital converters and specification points are the subject of Chap. 11.

For an overview of mathematical formulas on trigonometry and integral calculus some general reference books are: [6–8]. For statistical explanations have a look at [9, 10]. Device physics are explained in [11, 12] while [13] provides data on various technology options. Circuit design background is found in [14, 15].

Chapter 2
Sampling

2.1 Sampling in Time and Frequency

Sampling is a common technical process with the aim to represent a continuous-time signal by a sequence of samples. A movie consists out of a sequences of photographs (the samples), a newspaper photograph has been grated in little dots in two dimensions, a television broadcast consists out of a sequence of half pictures, etc. The sampling function can be implemented in many ways. In a photo camera the chemical substance on the film is exposed during the aperture time. In modern camera's the image sensor performs this function and allows light to generate charge during a short time. In all sampling realizations, a switch mechanism followed by some form of storage is required. In an electronic circuit a sample pulse defines the sampling moments and controls a switch (relays, bipolar, MOS, and avalanche device). There are two electronics storage media available: currents in a coil and voltages on a capacitor. The practical use of a switched coil is in the ignition circuit of a combustion engine, but is here outside of the scope. The most widely applied sampling circuit in microelectronics consists of a switch and a capacitor.

The analysis of the sampling process starts with a mathematical view. The sampling pulse determines the value of a signal on a predetermined frame of time moments. The sampling frequency f_s defines this frame and determines the sampling moments as:

$$t = \frac{n}{f_s} = nT_s, \quad n = -\infty, \ldots -3, -2, -1, 0, 1, 2, 3, \ldots \infty \tag{2.1}$$

Between the sampling moments there is a time period of length T_s, where mathematically speaking no value is defined.[1] In practice this time period is used to perform operations on the sample sequence. These various operations (summation,

[1] "No value is defined" does not imply that the value is zero! There is simply no value.

M. Pelgrom, *Analog-to-Digital Conversion*, DOI 10.1007/978-3-319-44971-5_2

multiplication, and delay) are described in the theory of time-discrete signal processing, e.g., [16, 17] and allow to implement filtering functions. In the present context the value of T_s is considered constant, resulting in a uniform sampling pattern. Generalized non-uniform sampling theory requires extensive mathematical tools and the reader is referred to the specialized literature.

Sampling transforms a time-continuous signal in a time-discrete signal and can be applied on all types of band-limited signals. In electronics, sampling of analog time-continuous signals into analog time-discrete signals is most common. Also time-continuous digital signals (like pulse-width modulated signals) and sampled signals themselves (as found in image sensors) can be (re-)sampled.

The mathematical description of the sampling process uses the "Dirac" function. This function $\delta(t)$ is a strange[2] mathematical construct as it is only defined within the context of an integral. The Dirac function requires that the result of the integral equals the value of the integral function at the position of the running variable as given by the Dirac function's argument.

$$\int_{t=-\infty}^{\infty} f(t)\delta(t-t_0)\,dt = f(t_0) \tag{2.2}$$

The dimension of the Dirac function is the inverse of the dimension of the running variable. A more popular, but not exact, description states that the integral over a Dirac function approximates the value "1":

$$\delta(t) = \begin{cases} 0, & -\infty < t \leq 0 \\ \frac{1}{\epsilon}, & 0 < t < \epsilon \\ 0, & \epsilon \leq t < \infty \end{cases} \Rightarrow \int_{t=-\infty}^{\infty} \delta(t)\,dt = 1 \tag{2.3}$$

with $\epsilon \to 0$.

A sequence of Dirac pulses mutually separated by a time period T_s defines the time frame needed for sampling:

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

This repetitive sequence of pulses with a mutual time spacing of T_s can be equated to a discrete Fourier series. The discrete Fourier transform (DFT) has sinusoidal frequency components with a base frequency $f_s = 1/T_s$ and repeats at all integer multiples of f_s. The amplitude factor for each frequency component at frequency kf_s is C_k. Equating both series yields:

$$\sum_{n=-\infty}^{n=\infty} \delta(t - nT_s) = \sum_{k=-\infty}^{\infty} C_k e^{jk2\pi f_s t} \tag{2.4}$$

[2]Strange in the sense that many normal mathematical operations cannot be performed, e.g., $\delta^2(t)$ does not exist.

As the Dirac sequence is periodic over one period T_s, the coefficients C_k of the resulting discrete Fourier series are found by multiplying the series with $e^{-jk2\pi f_s t}$ and integrating over one period.

$$C_k = \frac{1}{T_s} \int_{t=-T_s/2}^{T_s/2} \sum_{n=-\infty}^{n=\infty} \delta(t - nT_s) e^{-jk2\pi f_s t} dt \tag{2.5}$$

Within the integration interval there is only one Dirac pulse active at $t = 0$, so the complicated formula reduces to:

$$C_k = \frac{1}{T_s} \int_{t=-T_s/2}^{T_s/2} \delta(t) e^{-jk2\pi f_s t} dt = \frac{1}{T_s} e^{-jk2\pi f_s \times 0} = \frac{1}{T_s} \tag{2.6}$$

Now the substitution of C_k results in the mathematical description of the DFT from the sequence of Dirac pulses in the time domain.

$$\sum_{n=-\infty}^{n=\infty} \delta(t - nT_s) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} e^{jk2\pi f_s t} = \frac{1}{T_s} \left(1 + \sum_{k=1}^{\infty} 2\cos(k2\pi f_s t) \right) \tag{2.7}$$

Note that both terms are time-domain functions. The right-hand side is a summation of simple sine waves that can also be obtained from a frequency domain description using Dirac functions. This sum of Dirac functions in the discrete frequency domain is the counterpart of the time-domain Dirac sequence.

$$\sum_{n=-\infty}^{n=\infty} \delta(t - nT_s) \Leftrightarrow \sum_{k=-\infty}^{k=\infty} \delta(f - kf_s) \tag{2.8}$$

The infinite sequence of short time pulses corresponds to an infinite sequence of frequency components at integer multiples of the sampling rate.

2.1.1 Sampling Signals

In Fig. 2.1 a signal[3] $A(t)$ is sampled at a rate f_s. This signal corresponds in the frequency domain to $\mathbf{A}(f) = \mathbf{A}(\omega/2\pi)$ with a bandwidth from $f = 0$ Hz to $f = BW$. The straight-forward Fourier transform is defined as[4]:

[3] For clarity this chapter uses for time-domain signals normal print, while their spectral equivalents are in bold face. The suffix s refers to sampled sequences.

[4] The Fourier transform definition and its inverse require that a factor $1/2\pi$ is added somewhere. Physicists love symmetry and add $1/\sqrt{2\pi}$ in front of the transform and its inverse. Engineers mostly shift everything to the inverse transform, see Eq. 2.15. More attention is needed when a

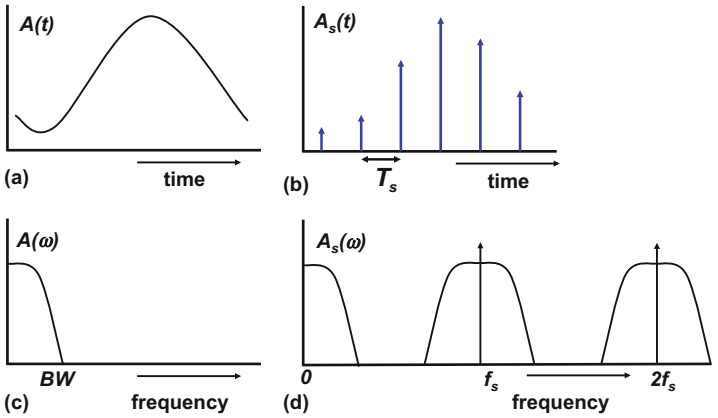


Fig. 2.1 Sampling an analog signal (**a**) in the time-continuous domain results in a series of analog signal samples (**b**). In the frequency domain the time-continuous signal (**c**) is folded around the sampling frequency and its multiples (**d**)

$$\mathbf{A}(f) = \int_{t=-\infty}^{\infty} A(t)e^{-j2\pi ft}dt \tag{2.9}$$

Mathematically sampling is performed by multiplying the time-continuous function $A(t)$ of Fig. 2.1a with the sequence of Dirac pulses, resulting in a time-discrete signal in Fig. 2.1b. The product of the time-continuous function and the Dirac sequence is defined for those time moments equal to the multiples of the sampling period T_s

$$A_s(t) = \int_{\tau=-\infty}^{\infty} A(t-\tau) \times \sum_{n=-\infty}^{n=\infty} \delta(\tau - nT_s)d\tau \tag{2.10}$$

A useful property of the Fourier transform is that a multiplication of time-domain functions corresponds to a convolution of their frequency counterparts in the Fourier domain.

$$\mathbf{A_s}(f) = \int_{\chi=-\infty}^{\infty} \mathbf{A}(f-\chi) \times \sum_{k=-\infty}^{k=\infty} \delta(\chi - kf_s)d\chi \tag{2.11}$$

Evaluation of this integral is easy as the result of an integral with a Dirac function is the integral function evaluated at the values where the Dirac function is active. In this case this means that for $k = 0$ the sampled data spectrum equals the original time-continuous spectrum: $\mathbf{A_s}(f, k = 0) = \mathbf{A}(f)$. For $k = \pm 1$ the result is $\mathbf{A_s}(f, k = \pm 1) = \mathbf{A}(f_s \pm f)$. Or in other words: the spectrum is mirrored around the first

single-sided engineering type Fourier transform is used instead of a mathematically more accurate double sided.

multiple of the sample rate. The same goes for $k = 2, 3, \ldots$ and negative k. Another property of the Fourier transform is applied: for a real function in the time domain, the Fourier result is symmetrical around 0: $\mathbf{A}(f) = \mathbf{A}(-f)$.

Consequently $\mathbf{A_s}(f)$ consists of a sum of copies of the time-continuous spectrum each with a frequency shift of $k \times f_s$. The total spectrum $\mathbf{A_s}$ can be written as:

$$\mathbf{A_s}(f) = \sum_{k=-\infty}^{\infty} \mathbf{A}(f - kf_s) \tag{2.12}$$

The original time-continuous signal $A(t)$ is connected to only one spectrum band in the frequency domain $\mathbf{A}(f)$. By sampling this signal with a sequence of Dirac pulses with a repetition rate (f_s) a number of replicas of the original spectral band $\mathbf{A}(f)$ are created on either side of each multiple of the sampling rate f_s. Figure 2.1c, d depicts the time-continuous signal and the sampled signal in the frequency domain. In the frequency domain of the sampled data signal, next to the original signal, also the upper bands are present.

The idea that from one spectrum an infinite set of spectra is created seems to contradict the law on the conservation of energy. If all spectra where mutually uncorrelated and could be converted in a physical quantity, there would indeed be a contradiction. However, in the reconstruction from a mathematical sequence of Dirac pulses to a physical quantity, there is an inevitable filtering operation, limiting the energy.

An important consequence of the previous sampling theory is that two frequency components in the time-continuous domain that have an equal frequency distance to arbitrary multiples of the sampling frequency will end up on the same frequency location in the sampled data band. Figure 2.2 shows three different sine wave signals that all result in the same sampled data signal (dots). Different signals in the time-continuous domain can have the same representation in the time-discrete domain.

A time-continuous signal close to $(m \times f_s)$ will result in replica around $((k \pm m) \times f_s)$. In case $k = m$ the signals around $(m \times f_s)$ will appear near DC, shifting the original signal band to low frequencies. This phenomena not only finds an application in down-mixing of communication signals, Sect. 2.3.2, but also means that unwanted or unexpected signals will follow the same path and show up in the wanted bandwidth. In the Sect. 2.2.2 countermeasures are discussed.

Example 2.1. The input stage of a sampling system is generating distortion. Plot the input signal of 3.3 MHz and its 4 harmonics in the band between 0 and a sampling frequency of 100 Ms/s. Change the sampling rate to 10 Ms/s and plot again.

Solution. Figure 2.3 shows on the left side the sampled spectrum at 100 Ms/s and on the right side the 10 Ms/s variant. Signal components will appear at frequencies: $i \times f_{in} \pm j \times f_s$, where $i = 1 \ldots$.number of harmonics and $j = 0 \ldots \infty$. Therefore in the last graph there are components at the frequencies shown in Table 2.1.

Note that within each $f_s/2$ range there are exactly five components, corresponding to each of the original tones.

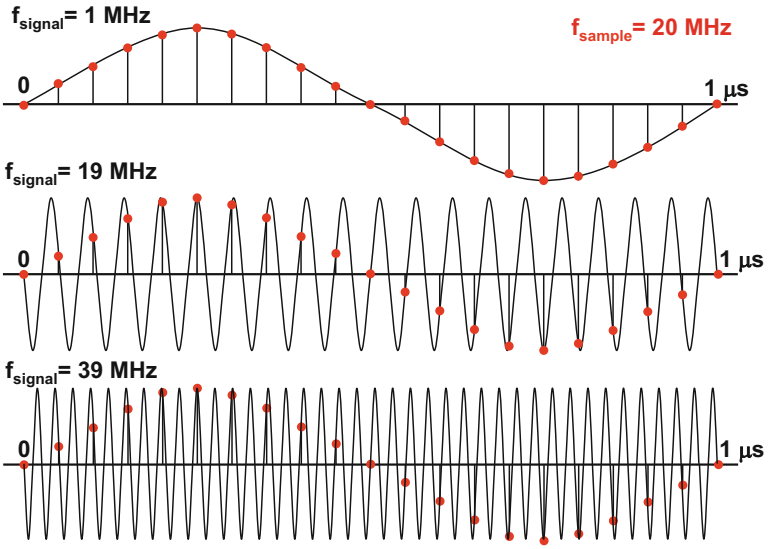


Fig. 2.2 Sampling three time-continuous signals: 1, 19, and 39 MHz sine waves result after sampling with 20 Ms/s in the same sampled data sequence (*dots*)

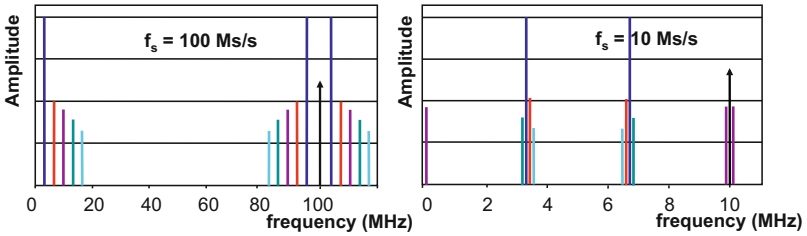


Fig. 2.3 The sample rate in the *left plot* is 100 Ms/s and in the right plot 10 Ms/s

Table 2.1 Frequency components of a distorted 3.3 MHz sinusoid sampled at 10 Ms/s

0.1 MHz	$f_s - 3f_{in}$
3.2 MHz	$f_s - 4f_{in}$
3.3 MHz	f_{in}
3.4 MHz	$f_s - 2f_{in}$
3.5 MHz	$f_s - 5f_{in}$
6.5 MHz	$2f_s - 5f_{in}$
6.6 MHz	$2f_{in}$
6.7 MHz	$f_s - f_{in}$
6.8 MHz	$2f_s - 4f_{in}$
9.9 MHz	$3f_{in}$

2.2 Sampling Limits

2.2.1 Nyquist Criterion

Figure 2.4 shows a signal in the time domain with higher frequency components than the signal in Fig. 2.1. The samples of this signal are valid values of the signal at that sample moments, however, it is not possible to reconstruct uniquely the signal based on these values. A likely reconstruction would be the dotted line, a signal that largely differs from the original.

If the bandwidth in the time-continuous domain increases, the mirror bands around the multiples of the sample frequency will follow. Figure 2.5 shows that this will lead to overlap of signal bandwidths after sampling, and mixing of data. This phenomenon is called "aliasing." The closest upper band directly adjacent to the original band is called: "the alias band." If the original signal mixes with its alias, its information contents are corrupted.

This limitation of sampling signals is known as the "Nyquist" criterion. Indicated already in a paper by Harry Nyquist [18, Appendix 2-a], it was Claude E. Shannon

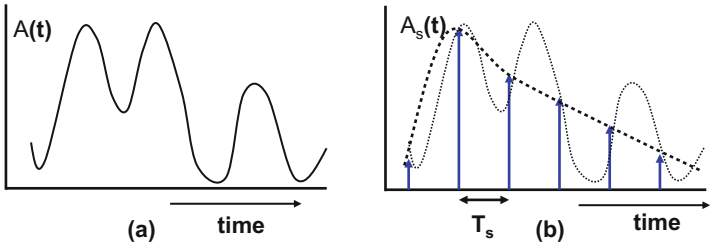


Fig. 2.4 The time-continuous signal contains higher frequency components (**a**) and does not satisfy the Nyquist criterion. The sample series in the time domain (**b**) allow multiple reconstructions of the original time-continuous signal

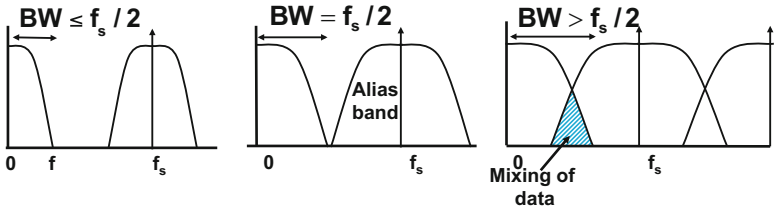


Fig. 2.5 The time-continuous bandwidth can be increased until half of the sample rate is reached

who extended his mathematical theory of communication [19] in 1949 with a paper dealing with communication in the presence of noise. In that paper [20] the Nyquist criterion, also known as Nyquist theorem, is formulated as[5]:

"If a function contains no frequencies higher than BW cycles per second, it is completely determined by giving its ordinates at a series of points spaced 1/2BW seconds apart."

This Nyquist criterion says that if the sample rate is more than twice the highest frequency in a bandwidth, there is a theoretical manner to uniquely reconstruct the signal. This criterion imposes a simple mathematical relation between a bandwidth BW and the minimum sample rate f_s:

$$f_s > 2BW \tag{2.13}$$

The Nyquist sample rate is often defined as $f_{s,ny} = 2 \times BW$ and the Nyquist bandwidth is the bandwidth $BW = f_{s,ny}/2$. The Nyquist frequency is the highest frequency in the Nyquist bandwidth.[6] This criterion is derived assuming ideal filters and an infinite time period to reconstruct the signal. In practical circumstances designers will use additional margins to avoid having to meet these constraints. An interesting discussion on present insights in the mathematical aspects of the Nyquist criterion was published by Unser [21].

The Nyquist criterion specifies that the useable bandwidth is limited to half of the sample rate. But the Nyquist criterion does not define where the bandwidth is located in the time-continuous spectrum. The only constraint on the position of this limited bandwidth that this bandwidth does not include any multiple of half of the sample rate. That would lead to overlap in the sampled spectrum. There is no need to specify the bandwidth starting at 0 Hz. For example, if it is known that the original signal in Fig. 2.2 is in the bandwidth between 10 and 20 MHz, the samples can be reconstructed to yield the originating 19 MHz time-continuous sine wave. A bandwidth, located in the spectrum at a higher frequency than the sample rate, can therefore also be properly sampled. The sampling operation generates copies around all multiples of the sample rate, including near DC. This is in some communication systems used to down-modulate or down-sample signals, see Sect. 2.3.1.

Example 2.2. A 10 kHz sine wave is distorted with components at 20, 30, 40, and 50 kHz, and sampled with 44 ks/s. Draw the spectrum.

Solution. In the left part of Fig. 2.6 shows the input spectrum. The right part shows the result after sampling. The tones from the original spectrum are in bold lines, the results of the folding and mirroring around the 44 ks/s sample rate are drawn

[5] Other originators for this criterion are named as E. Whittaker and V.A. Kotelnikov. Landau proved in 1967 for non-baseband and non-uniform sampling that the average sample rate must be twice the occupied bandwidth.

[6] Precise mathematicians will now argue that a signal with frequency $f = f_{s,ny}/2$ cannot be reconstructed, so they should read here: $f = f_{s,ny}/2 - \Delta f$, where Δf goes to zero.

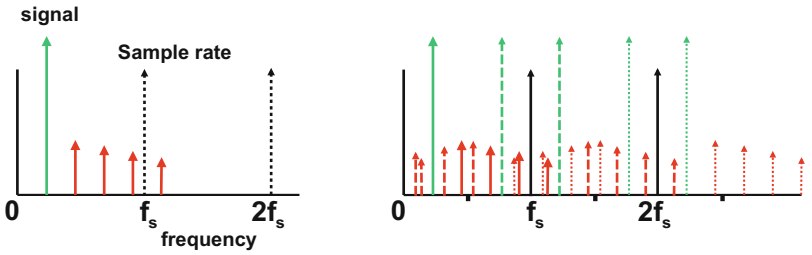


Fig. 2.6 The sample rate in the upper plot is 44 ks/s

in dashed lines and the components originating from the second multiple at 88 ks/s are shown in thinner dashed lines. The 50 kHz component results after sampling in a (−6) kHz frequency. This component is shifted to the positive frequency domain, while keeping in mind that it differs in phase. In every interval of $f_s/2$ width there is exactly one copy of each originating component. So there is a simple check on the correctness and completeness of the spectrum: make sure the number of components exactly matching the number at the input.

Note here, that the discrete tones as they are, have zero bandwidth. And as long as folding of one tone on top of another is prevented, or is considered irrelevant, many forms of sampling can be applied.

Example 2.3. A sampling system with distortion is excited with a single-tone sine wave. An unexpected tone appears in the output spectrum. How do you find out what its origin is?

Solution. In a sampled data system it is obvious that tones can be generated by distortion of the sine wave in combination with aliasing through sampling. The first check is made by varying the input frequency by a small frequency offset Δf_{in}. If the tone in the output spectrum varies by a multiple of the offset $i \times \Delta f_{in}$, the i-th harmonic of the input tone is involved. Now the same procedure is repeated for the sampling rate, identifying $j \times \Delta f_s$ as the originator. The integers i, j in the formula $i \times f_{in} \pm j \times f_s$ are known and the evaluation of this formula should point at the unexpected tone position.

There are of course many more possible scenarios, e.g., suppose there is a reaction on varying the sampling rate, but not on input signal variations. Probably an external frequency is entering the system and gets mixed down in the sampling chain.

2.2.2 Alias Filter

The Nyquist criterion requires that all input signals are band-limited in order to prevent mixing up of modulated signal components. This requirement is even more

stringent for signals that are up-front unwanted: various noise contributions, tones, distortion, etc. These signals are also modulated by the (integer multiples of) the sample rate. As this sampling process stretches out into high-frequency ranges, even RF signals can cause low frequency disturbance after modulation with hundreds $\times f_s$. In a correctly designed analog-to-digital conversion system the bandwidth of the incoming signal is limited by means of an "alias-filter," so that no mixing of out-of-band frequency components can take place, see Fig. 2.7. An analog-to-digital converter is therefore preceded by a band-limiting filter, see Fig. 2.8. This filter prevents the components outside the desired frequency range to be sampled and to mix up with the wanted signals.

In practical system design it is recommended to choose a higher sample rate than prescribed by the Nyquist criterion. The fraction of frequency spacing between the extremes of the base and its alias with respect to the sample rate determines the number of poles needed in the anti-alias filter. A filter will suppress signals at a rate of 6 dB per octave per filter pole, Fig. 2.9.

Sharp band-limiting filters require many accurately tuned poles. Additional amplification is needed, and therefore these filters tend to become expensive and hard-to-handle in a production environment. On the other hand, there are some good reasons not to choose for an arbitrary high sample rate: the required capacity for storing the digital data will increase linear with the sample rate, as well as the power needed for any subsequent data processing.

Anti-alias filters are active or passive time-continuous filters. Time-discrete filters, such as switched-capacitor filters, sample the signal themselves and require consequently some alias filters. An additional function of the anti-alias filter can be the suppression of unpredictable interference in the system. Note that interference

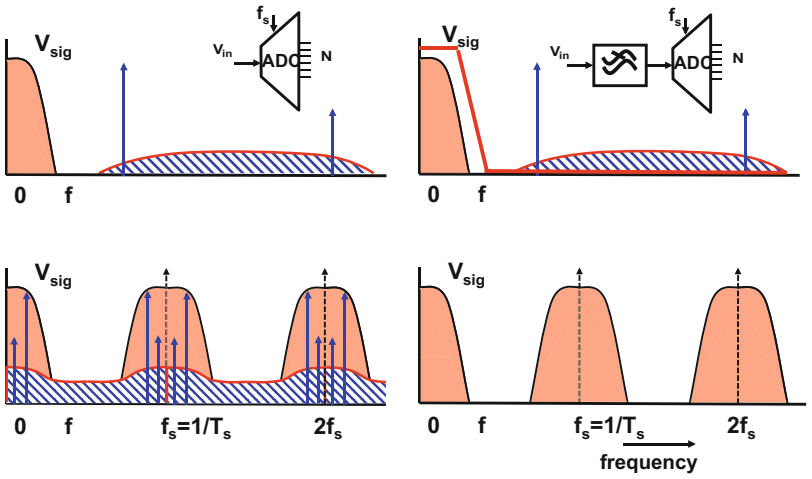


Fig. 2.7 *Left*: sampling of an unfiltered signal leads to lots of unwanted components in the signal. *Right*: an alias filter prevents disturbing signals, tones, or spurs to enter the signal band

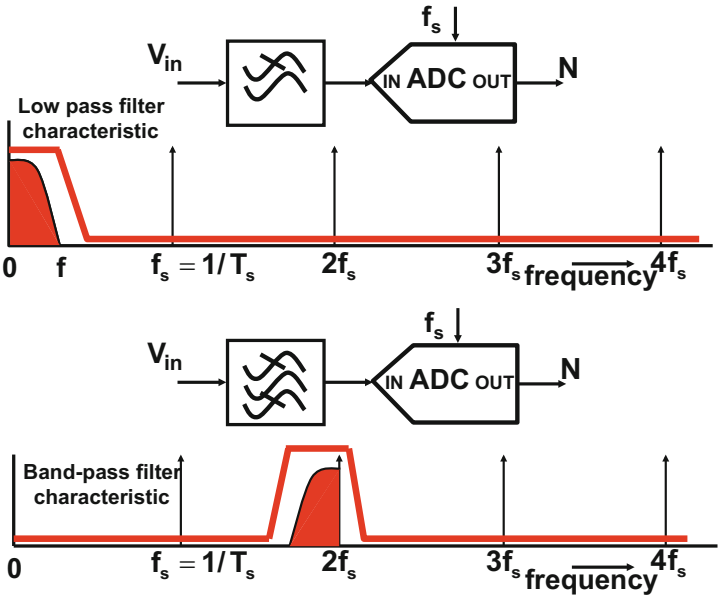


Fig. 2.8 A low-pass filter (*upper*) or bandpass filter (*lower*) is used to avoid unwanted components near other instances of the sample rate

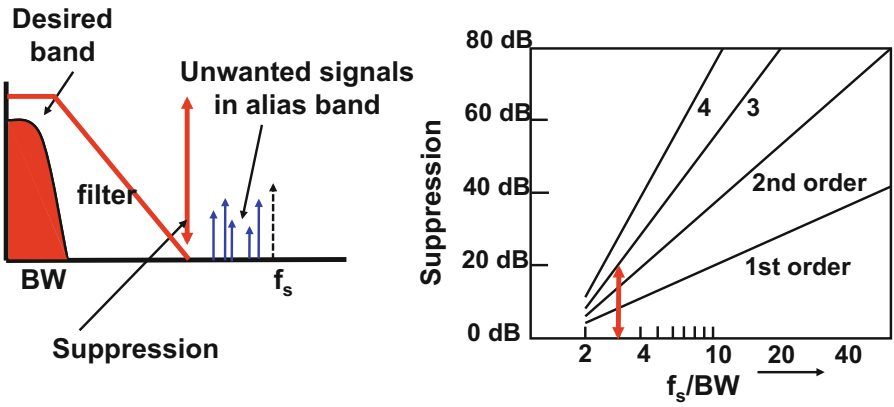


Fig. 2.9 The attainable suppression of the anti-alias filter depends on the number of poles in the filter and the ratio of the bandwidth to the sample rate

does not necessarily enter the system via its input terminal, the designer should have an equal interest in suppressing any interference on supply and bias lines.

Some systems are band-limited by construction. In a radio, the IF filters of a heterodyne radio architecture may serve as anti-alias filters, and in a sensor system, the sensor may be band-limited.

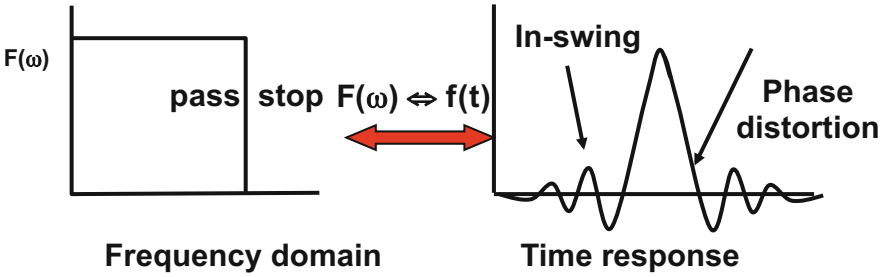


Fig. 2.10 The mathematical description of the relation between frequency domain and time domain implies that a sharp-limited frequency response generates a ringing response in the time domain

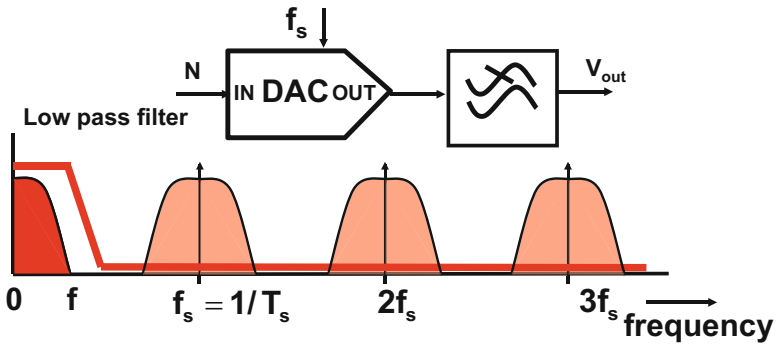


Fig. 2.11 Before using the result of the reconstruction the higher harmonic bands must be removed. In this example it is assumed that the reconstruction provides no filtering

The definition of a filter for alias suppression requires to look at pass-band, rejection but also at signal ringing. Figure 2.10 shows a brick-wall filter characteristic.

$$\mathbf{F}(\omega) = \begin{cases} 0, \ \omega \leq 0 \\ 1, \ 0 < \omega \leq \omega_{BW} \\ 0, \ \omega > \omega_{BW} \end{cases} \tag{2.14}$$

$$f(t) = \frac{1}{2\pi} \int_{\omega=-\infty}^{\infty} \mathbf{F}(\omega) e^{j\omega t} d\omega \tag{2.15}$$

$$f(t) = \frac{2\sin(\omega_{BW} t)}{2\pi t} \tag{2.16}$$

The impulse response is a $\sin(x)/x$ function with ringings on both sides of the pulse. These ringings will be triggered by transitions in the signal and are disastrous in many applications. On a television screen a vertical stripe would on either side be accompanied by shadows. In an audio system these ringings lead to phase distortion, or in more musical terms: blurring of the instrument's position. These practical considerations often lead to a sample rate of at least 2.5–3× the bandwidth.

Alias filtering at the input of a conversion chain is necessary to remove unwanted components in the spectrum that may fold back into the signal after sampling. Also at the output side of the conversion chain an alias filter can be necessary as the sampled data format contains high-frequency components, Fig. 2.11. During reconstruction, see Sect. 2.4, some filtering will occur, but mostly additional filtering is needed to avoid problems. If these components are processed in a non-linear fashion, unwanted signals can be produced. Also other failures may occur. For example, in an audio chain, the speakers are dimensioned assuming that (by far) most audio energy is in the low frequency range. If too much alias products exist, the tweeters can be harmed.

Example 2.4. A bandwidth of 2 MHz is sampled at 10 MHz. Determine the order of an anti-alias filter build as a cascade of equal first order filters, suppressing alias components by 35 dB.

Solution. Aliasing will occur due to signals in bands around multiples of the sampling rate. In this case all signals between $f_s - BW = 8$ and $f_s + BW = 12$ MHz will appear after sampling in the desired signal band. The task will be to design a filter that passes a bandwidth of 2 MHz, but suppresses at 8 MHz. The transfer expression is

$$H(\omega) = \left| \frac{1}{1 + (\omega\tau)^2} \right|^{n/2}$$

where n is the filter order and assuming that $\omega\tau = 1$ for a frequency of 2 MHz, then a 3th order filter is chosen. This filter will attenuate the signal at 2 MHz by 9 dB. If only 3 dB attenuation is allowed, the filter order must be increased to 7. It is obvious that doubling the sample rate eases this trade-off dramatically.

Example 2.5. Comment on the choice of the sample rate in the CD audio standard.

Solution. An example of a critical alias filter is found in the compact-disc music recording format. Here a sample rate of 44.1 ks/s[7] is used for a desired signal bandwidth of 20 kHz. This combination leaves only a small transition band between 20 and 24.1 kHz to suppress the alias band by some 90 dB. The expensive filter required to achieve this suppression needs some 11–13 poles. Moreover such a filter will create a non-linear phase behavior at the high baseband frequencies. Phase distortions are time distortions (Δphase equals signal frequency $\times\Delta$time) and have a strong audible effect. Fortunately the use of "oversampling" allows to separate baseband and alias band sufficiently, see Sect. 10.1.

[7]The only storage in the early days of CDs were video recorders. The 44.1 ks/s sample rate was chosen such that the audio signal exactly fits to a video recorder format (25 fields of 588 lines with 3 samples per line) of 44.1 ks/s.

2.2.3 Getting Around Nyquist?

An implicit assumption for the Nyquist criterion is that the bandwidth of interest is filled with relevant information. This is not necessarily true in all systems. In communication systems like Wifi or GSM, only a few channels in the allocated band will be used at any moment in time. Video signals are by their nature sampled signals: a sequence of images consisting of sequences of lines. The spectral energies are concentrated around multiples of the video line frequency. The intermediate frequency bands are empty. These systems show "sparsity" in the frequency domain.

In a radar or ultra-sound system a pulse is generated and transmitted. The only relevant information for the system is the moment the reflected pulse is received. This is an example of time sparsity.

A sparse signal in a relatively wide bandwidth can be reconstructed after sampling by a non-uniform sampling sequence. Such a sequence can be generated by a high-frequency random generator. The information from the few active carriers is spread out over the band and theoretically it is possible to design algorithms that recover this information. A first intuitive approach is to assume a high uniform sampling pattern, from which only a few selected samples are used. In a higher sense the Nyquist criterion is still valid: the total amount of relevant bandwidth (Landau bandwidth) is still less than half of the effective sample rate.

Compressive sensing or compressive sampling [22] multiplies the signal with a high-rate random sequence, which is easier to implement in the analog domain than sampling. The relevant signals are again spread out over a large bandwidth. After bandwidth-limiting, a reconstruction ("L1" minimization) is possible if the random sequence is known and the domain in which the signal is monitored, is sparse. The theory is promising but requires heavy post-processing. Whether a real advantage can be obtained remains to be proven.[8]

Example 2.6. Two sine wave signals at 3.2 and 4.8 MHz each modulated with a 0.1 MHz bandwidth signal are sampled at 1.1 Ms/s. Is the Nyquist criterion violated?

Solution. No, the total band occupied with relevant signal is 0.4 MHz, while the Nyquist bandwidth is 0.55 MHz. The sample rate must be carefully chosen not to mix things. Here the sampled bandwidths will span 0–0.2 MHz and 0.3–0.5 MHz.

[8] Keep track of: dsp.rice.edu/cs for an overview of all compressive sampling developments.

2.3 Modulation and Chopping

2.3.1 Modulation

Sampling of signals resembles modulation of signals. In both cases the operation results in the creation of frequency shifted bands of the original signal. A modulator multiplies the baseband signal with a sine wave, resulting in a pair of upper bands around the carrier frequency, see Fig. 2.12. Mathematically modulation is the multiplication of a signal with a pure sine wave of radial frequency ω_{local}:

$$G_{mix}(t) = A(t) \times \sin(\omega_{local} t) \tag{2.17}$$

In the simple case of $A(t) = A\sin(\omega t)$:

$$A\sin(\omega t) \times \sin(\omega_{local} t) = \frac{A}{2}\cos((\omega_{local} - \omega)t) - \frac{A}{2}\cos((\omega_{local} + \omega)t) \tag{2.18}$$

In the result there are no components left at the input frequencies. Only two distinct frequencies remain. This modulation technique is the basis for the first radio transmission standard: amplitude modulation.

If A(t) is a band-limited spectrum composed of many sinusoidal signals, mixing results in two frequency bands:

$$G_{mix}(t) = A(t) \times \sin(\omega_{local} t)$$
$$\mathbf{G_{mix}}(\omega) = \frac{1}{2}\mathbf{A}((\omega_{local} - \omega) - \frac{1}{2}\mathbf{A}(\omega_{local} + \omega)$$

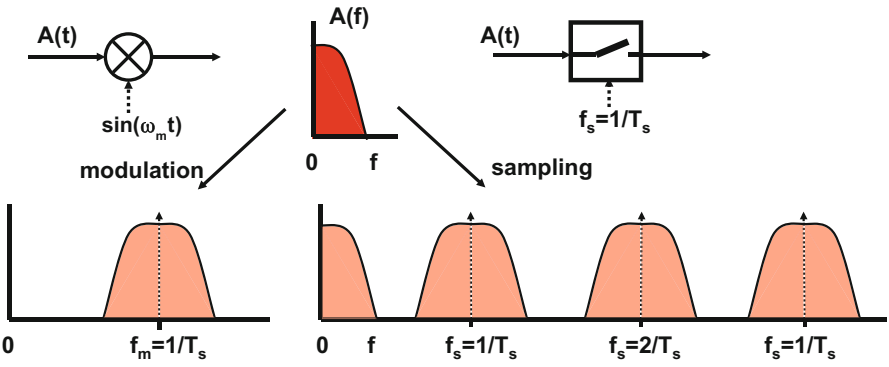


Fig. 2.12 Modulation and sampling of signals. Ideally the modulation and sampling frequencies disappear from the resulting spectrum. Here they are indicated for reference

The modulated bands appear as mirrored copies of each other around ω_{local}. Often one band is desired and the other band is referred to as the "mirror image."

If the modulation principle is repeated the original sine wave can be recovered:

$$G_{mix-down}(t) = G_{mix}(t) \times \sin(\omega_{local}t)$$

$$\left(\frac{A}{2}\cos((\omega_{local}-\omega)t) - \frac{A}{2}\cos((\omega_{local}+\omega)t)\right) \times \sin(\omega_{local}t) =$$

$$\frac{A}{2}\sin(\omega t) - \frac{A}{4}\sin(2\omega_{local}t+\omega t) + \frac{A}{4}\sin(2\omega_{local}t-\omega t) \tag{2.19}$$

The original component is accompanied by a pair of frequencies around $2\omega_{local}$. With a low-pass filter these components are removed.

In contrast to modulation, sampling results in upper bands around every multiple of the sample rate. The sequence of Dirac pulses is equivalent to a summation of sine waves with frequencies at multiples of the sample rate.

$$\text{Time domain} \sum_{n=-\infty}^{n=\infty} \delta(t-nT_s) = \frac{1}{T_s}\sum_{k=-\infty}^{\infty} e^{jk2\pi f_s t} = \frac{1}{T_s} + \frac{2}{T_s}\sum_{k=1}^{\infty}\cos(k2\pi f_s t)$$

$$\text{Frequency domain} \quad \mathbf{D_s}(\omega) = \frac{2\pi}{T_s}\sum_{k=-\infty}^{k=\infty}\delta(\omega - \tfrac{2\pi k}{T_s}) \tag{2.20}$$

This sequence of Dirac functions in the frequency domain translates back to sine waves in the time domain with frequencies that are integer multiples of the sample rate. Therefore sampling can be viewed as a summation of modulations. The intrinsic similarity between sampling and modulation is used in various system architectures: an example is found in down-mixing of radio frequency signals.

A particular aspect of sampling and mixing is called "self-mixing." A mixer can be seen as a device with two (mathematically) equivalent input ports. If a fraction of the signal on one port leaks into the other port, self-mixing will occur. If this leakage is described as $\alpha\sin(\omega_{local}t)$, the resulting output component will contain terms of the form: $\alpha/2 + \sin(2\omega_{local}t)/2$. In practical circuits mostly the large-amplitude local-oscillator frequency will leak into the low-amplitude port or the sample frequency is injected on the input node. A noticeable DC component is the result that can easily be mistaken for a circuit offset.

2.3.2 Down-Sampling

In the description of signals in the previous paragraphs, implicitly the band of interest was assumed to be a baseband signal, starting at 0 Hz with a bandwidth $f = BW$. This is the situation that exists in most data-acquisition systems. The mirror bands will appear around the sample rate and its harmonics. The choice for

this location of the band of interest is by no means obligatory and certainly not imposed by the Nyquist criterion. A band of interest located at a higher frequency, or even beyond the sample rate, can be sampled equally well. The signal band can be regarded as being sampled by the closest multiple of the sample rate. This band is again copied to all integer multiples of the sample rate, including "0 Hz". This process is called "under-sampling" or "down-sampling."[9] If there are components of the signal lying on equal frequency spacings above and below a multiple of the sample rate, both of these will be sampled into the same frequency region. The consequence is an overlap of signals and must be avoided.

Deliberate forms of down-sampling are used in radio-communication applications, where down-sampling is used as a way to perform demodulation, see Fig. 2.13.

Unwanted forms of down-sampling occur if undesired signals are present in the signal band. Examples are

- Harmonic distortion products of the baseband signal.
- Thermal noise in the entire input band, see Sect. 2.5.1.
- Interference signals from other parts of the equipment or antenna.

Example 2.7. Set up a down-sampling scheme for an FM-radio receiver.

Solution. In an FM-radio the IF signal of 100 kHz is located at a carrier frequency of 10.7 MHz. This signal can be down-modulated and sampled at the same time by a 5.35 Ms/s signal, Fig. 2.13.

Essentially any sample rate that fulfills $f_s = 10.7/i$ Ms/s, where i is a positive integer, will convert the modulated band from 10.7 MHz to DC.

Two dominant considerations play a role in the choice of the sample rate. A low sample rate results in a low power consumption of the succeeding digital circuitry and less memory if storage is needed. A high sample rate creates a wide empty frequency range and allow easy and cheap alias filtering. Sometimes the sample rate

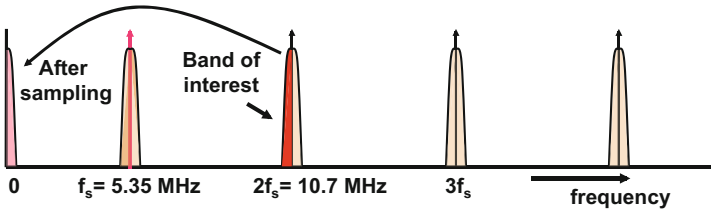


Fig. 2.13 Demodulation and down-sampling of an IF-FM signal at 10.7 MHz by a 5.35 Ms/s sample pulse

[9]In this book the term down-sampling is used for sampling an analog signal with the purpose to perform a frequency shift of the band of interest. Subsampling in Sect. 2.3.3 removes samples in a predetermined manner from an existing sample stream, but does not change the signal band.

can be chosen in such a manner that undesired input frequencies end up in an unused part of $f = 0, \ldots, f_s/2$.

Example 2.8. An IF-television signal occupies a bandwidth from 33 to 39 MHz. Propose a sampling frequency that maps the 39 MHz component on DC. Consider that the power consumption of the following digital circuit is proportional to the sampling frequency and must be low.

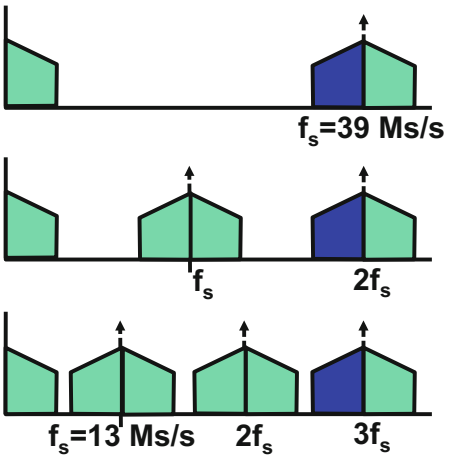


Fig. 2.14 Three solutions for sampling a bandwidth between 33 and 39 MHz

Solution. A sampling rate of 78 Ms/s misses the point in the Nyquist criterion, as the bandwidth is only 6 MHz and not 39 MHz. The Nyquist rate is 12 Ms/s so alternatives are shown in Fig. 2.14. A sample rate of 39 Ms/s will work, but causes a lot of digital power.

A sample rate of 19.5 Ms/s is a viable alternative to 39 Ms/s as it halves the digital power but leaves enough frequency space for alias filtering.

And a sample rate of 13 Ms/s which will leave only a 1 MHz frequency range for alias filtering. This might be an expensive solution when the alias has to be suppressed.

2.3.3 Subsampling and Decimation

In some applications reducing the sample rate is a necessity. In sigma-delta conversion, decimation or subsampling is a necessity to translate the high-speed bit-stream signal in normal samples. Another example is in measuring the performance of a very high-speed sample system. This requires a test-setup with even better specifications at those frequencies. Here too, subsampling reduces the sampling frequency and allows to measure accurately, without the need for extreme performance.

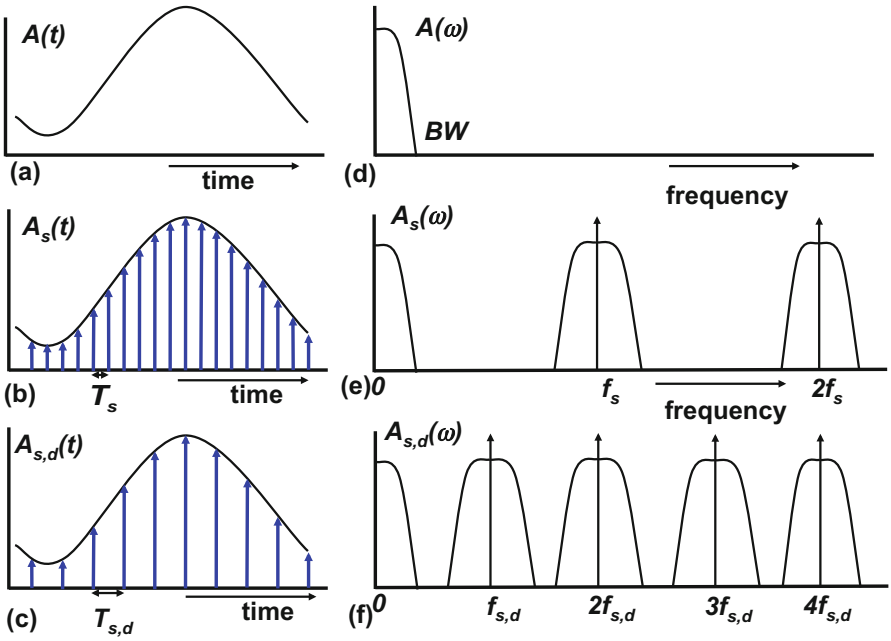


Fig. 2.15 Decimation or subsampling of a sampled signal. *Upper*: a time-continuous signal in the time (*left*) and frequency domain (*right*). *Middle*: Time and frequency representation after sampling. *Lower*: Time and frequency representation after subsampling by a factor of two

Figure 2.15 shows the basic process of subsampling or decimation. In Fig. 2.15a, b, d, e the time sequence and frequency representation of a signal sampled at f_s are shown. If this signal is subsampled by an integer factor[10] M (in this figure $M = 2$) every M-th sample is kept and the unused samples are removed, see Fig. 2.15c, f.

The procedure in Fig. 2.15 can be used because the bandwidth of the original signal is less than half of the new sample rate. Thereby this signal fulfills the Nyquist criterion for the new sample rate f_s/M. In cases where this is not the case, the bandwidth must be sufficiently reduced, before subsampling is applied. If this is omitted, serious aliasing will occur with loss of information.

2.3.4 Chopping

Chopping is a technique used for improving accuracy by modulating error-sensitive signals to frequency bands where the signal processing is free of errors, see also

[10]Subsampling by a rational factor (a division of two integers) or an irrational factor requires to calculate the signal at each new sample moment by interpolation of the existing samples. This technique is often applied in image processing and is used to combine data from sources with asynchronous clocks. Some fast-running hardware is needed to carry out the interpolation.

Sect. 7.6. In Fig. 2.16 first the signal is modulated to a higher frequency band by multiplication with a chopping signal $f_{chop}(t)$. After signal processing, the signal is modulated back by multiplying again with $f_{chop}(t)$. The technique works well with a sine wave or a block wave as modulator as $f^2_{chop}(t)$ contains a DC-term and for the rest only frequency components far above the band of interest. Chopping can also be used to move unwanted signals out of the band of interest. For example, alternating between DC-current sources (also known as dynamic element matching) will move mismatch and the $1/f$ noise to higher bands.

In differential circuits, chopping is implemented easily by alternating between the differential branches. Mathematically this correspondents to a multiplication with a block wave with amplitude $+1, -1$. This block wave can be decomposed into a series of sine waves:

$$f_{chop}(t) = \sum_{n=1,3,5,..}^{\infty} \frac{4\sin(n\pi/2)}{n\pi} \cos(\omega_{chop} t) \tag{2.21}$$

Now $f^2_{chop}(t) = 1$ and a perfect restoration after chopping back is possible. Note that a signal $f_{chop}(t)$ composed of any sequence of $+1, -1$ transitions, at fixed frequency or at arbitrary time moments, shows this property and can be used for chopping purposes. As Fig. 2.17 shows, chopping with a block wave can be done with lower frequencies than the bandwidth of the input signal. Chopping does not compress the signal into one single value, and consequently there is no direct impact of the Nyquist criterion on chopping. On the other hand, chopping is a form of modulation and so any unwanted signals entering the chopping chain between the two modulators may cause problems and alias filtering is a remedy.

The spectrum of a block wave fixed-frequency chopped signal will be composed of a series of modulated spectra around odd multiples of the chopping frequency:

$$\mathbf{A_{chop}}(\omega) = \sum_{n=1,3,5,..}^{\infty} \frac{4\sin(n\pi/2)}{n\pi} \mathbf{A}(n\omega_{chop} \pm \omega)$$

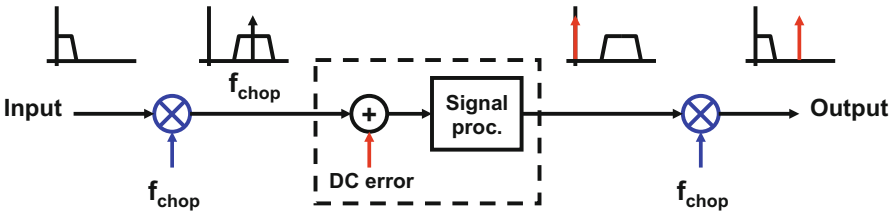


Fig. 2.16 A simple chopping chain: the input signal must be protected against the unwanted DC-term. Two chopping modulators move the signal band to a high frequency and back, thereby avoiding the DC term. This configuration was a standard technique in precision electronics of the 1950s

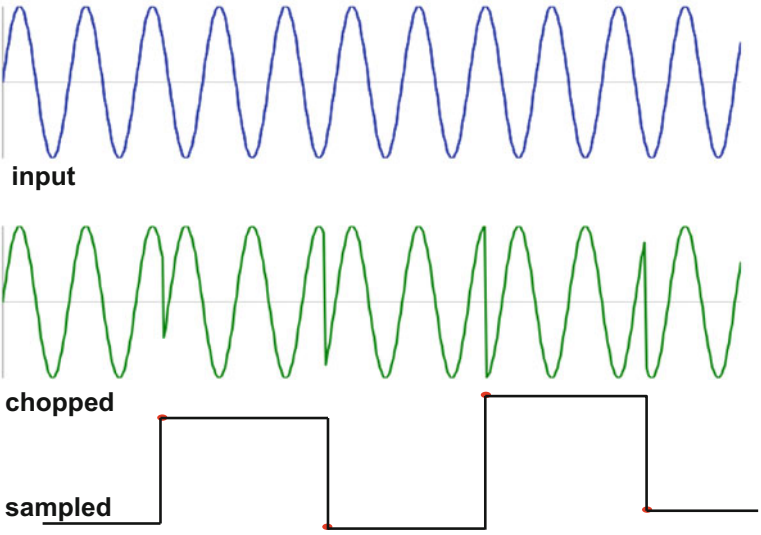


Fig. 2.17 Chopping (*middle*) and sampling (*below*) differ fundamentally in their information contents

E.g. chopping a spectrum from 0 to 1 MHz with a block wave $(+1, -1)$ of 10 MHz, will remove the spectrum near DC and generate mirror bands at 9–11, 29–31, 49–51 MHz, etc.

The higher bands of the chopped signal should not be removed. This would cause imperfections after chopping back. Any removed components can be regarded as a negative addition of signals to a perfectly chopped spectrum. These components will be treated as new input signals for the chopping back operation. So products of these components with the signal of Eq. 2.21 will appear. The removed parts of the spectrum $\mathbf{A}(n\omega_{chop} \pm \omega)$ will be modulated by the n-th harmonic of Eq. 2.21, resulting is an amplitude contribution at the position of the original signal with a relative strength of $1/n^2$.

Example 2.9. A 135 MHz sine wave is sampled in a 150 Ms/s sampling system. Which frequency components will be in the sampled data spectrum? Is it possible to discriminate the result of this sampling process from sampling a 15 MHz sine wave?

Solution. If an input signal at frequency f_i is sampled by a sample rate f_s than the sample data spectrum will contain the lowest frequency from the series: $f_i, (f_s - f_i), (f_s + f_i), (2f_s - f_i), (2f_s + f_i), \ldots (nf_s - f_i), (nf_s + f_i), \ldots$ where $n = 0, 1, 2, \ldots .\infty$. In this case the second term delivers a 15 MHz component.

If directly a 15 MHz sine wave was sampled the sampled data sequence would be similar and even perfectly identical provided that the mutual phase shift is correct. In perfect conditions there is no way to tell from which time-continuous signal (in this case 15 or 135 MHz) this sequence originates. Continued in Example 2.16 on page 40.

2.4 Reconstruction of Sampled Data

The sequence of samples (after analog-to-digital conversion and any form of digital signal processing) that arrives at the input of a digital-to-analog converter is a set of numerical values corresponding to the frame of sample moments. A spectral analysis would result in an ideal sampled data spectrum, where all copies of the signal band at multiples of the sample rate are equivalent. In the time domain the value of the signal between the sample moments is (mathematically spoken) not defined.

This stream of samples must at some instant be reconverted in the time-continuous domain. The first question is what to do with the lacking definition of a signal in between the samples. The most common implementation to deal with this problem is to simply use the value of the signal at the sample moment and to keep it for the entire sample period. Figure 2.18 (left) shows this "zero-order hold" mode. A more sophisticated mechanism interpolates between two values as in Fig. 2.18 (middle). An elegant form of interpolation uses higher-order or spline-fit algorithms, Fig. 2.18 (right).

In most digital-to-analog converters a zero-order hold function is sufficient because the succeeding analog filters perform the interpolation. Moreover a zero-order hold operation is often for free as the digital input signal is stored during the sample period in a set of data latches. The conversion mechanism (ladders or current sources) simply converts at any moment whatever value the data latches hold. This option has several additional advantages. Whenever the output signal of a digital-to-analog converter contains glitches, an explicit sample-and-hold circuit will remove the glitches and improve the quality of the conversion. In case of algorithmic digital-to-analog converters the output signal has to be constructed during the sampling period (see, e.g., Sect. 7.4.4). Then a hold circuit shields any incomplete conversion results and prevents them to appear at the output.

Holding the signal during a period $T_h \leq T_s$ changes the shape of the signals passing through a zero-order hold operation. Holding of the signal creates a signal transfer function. The impulse response of the hold transfer function is found by considering that the Dirac sequence is multiplied by a function consisting of a constant term "1" over the hold period T_h.

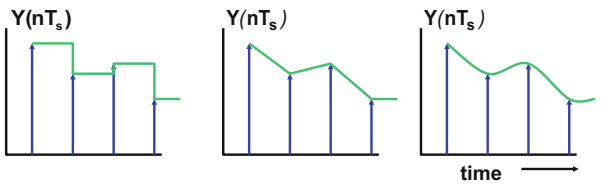


Fig. 2.18 A Dirac sequence can be reconstructed via a zero-order hold (*left*), a first-order interpolation (*middle*) or higher-order reconstruction algorithms (*right*)

$$h(t) = \begin{cases} 1, & 0 < t < T_h \\ 0, & \text{elsewhere} \end{cases} \tag{2.22}$$

The frequency domain transfer function $H(\omega)$ of a zero-order hold function (implemented in Chap. 3 as a sample-and-hold circuit) is calculated via the Fourier transform. The result of this transform has the dimension time.[11] In order to obtain a dimensionless transfer function, a normalization to T_s is introduced:

$$H(\omega) = \int_{t=0}^{t=\infty} h(t) \times e^{-j\omega t} dt = \int_{t=0}^{t=T_h} 1 \times e^{-j\omega t} dt = \frac{\sin(\pi f T_h)}{\pi f} e^{-j\omega T_h/2} \Leftrightarrow \frac{\sin(\pi f T_h)}{\pi f T_s} e^{-j\omega T_h/2} \tag{2.23}$$

Figure 2.19 shows the time and frequency response of the zero-order hold function for various values of the hold time T_h. The mathematical formulation of the amplitude function is often summarized to "$\sin(x)/x$" behavior. Some authors use "$\mathrm{sinc}(x)$". The integral of the function $\sin(x)/x$ belongs to the mathematical class of Dirichlet integrals, with as property:

$$\int_{x=0}^{\infty} \frac{\sin(x)}{x} dx = \pi/2 \tag{2.24}$$

The last term in Eq. 2.23 is $e^{-j\omega T_h/2}$ which represents a delay in the time domain. This delay $T_h/2$ is introduced as the value of the signal that was first concentrated in the sample moment is now distributed over the entire hold period. The average value moves from the sampling moment (the edge of the clock pulse) to the middle of the hold period.

A zero response occurs at frequencies equal to multiples of the inverse of the hold time. Obviously signals at those frequencies complete one or more complete periods in the hold time and exactly average out. For short hold periods approximating a Dirac function, this zero moves to infinity and the transfer of the sample-and-hold circuit is flat over a large frequency range. If T_h becomes equal to the sample period T_s the transfer function shows a zero at the sample rate and its multiples.

The amplitude response in the frequency domain is a representation of the average energy over (theoretically) infinite time. In the time domain sample values and consequently zero-order hold values can occur with amplitudes equal to the maximum analog input amplitude. A signal close to half of the sampling rate can show in one time period a small amplitude while achieving a value close to the full range input signal at another time instance depending on the phase relation of the signal and the sample rate. Still this signal has over infinite time an averaged

[11] Formally the result of a Fourier transform reflects the intensity of a process or signal at a frequency. Therefore the result has the dimension "events per Hz" or "Volt per Hertz."

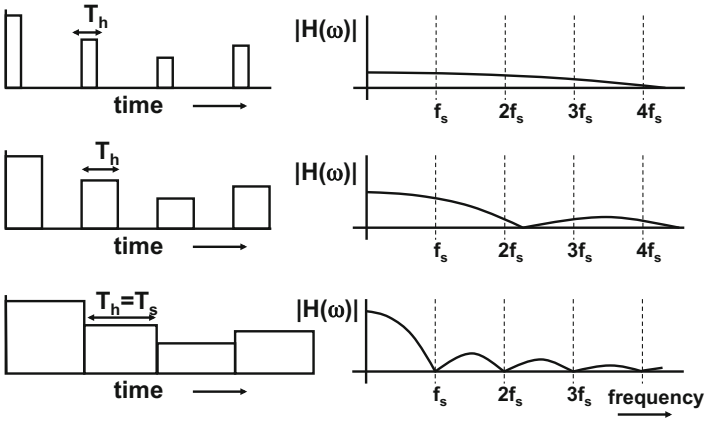


Fig. 2.19 The hold time determines the filter characteristics of a sample-and-hold function

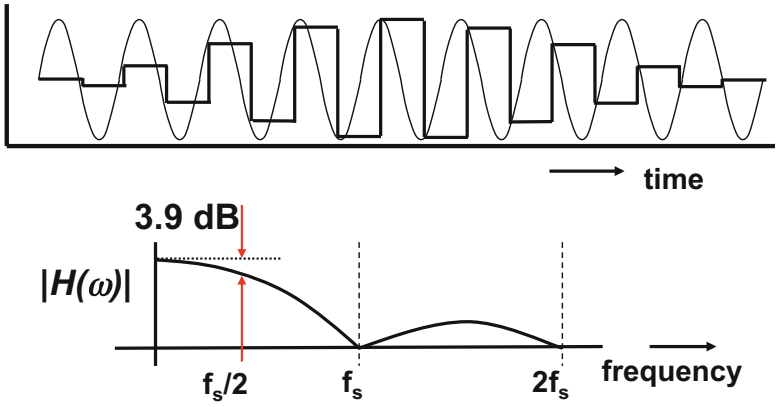


Fig. 2.20 Time and frequency response of a sample-and-hold signal close to half of the sampling rate

attenuation of 3.9 dB. In that sense the attenuation in Fig. 2.20 is different from a frequency transfer function of, e.g., an R-C network, where the attenuation at a certain frequency is independent of the phase.

Example 2.10. Can the $sin(x)/x$ frequency behavior of a zero-order hold circuit be compensated by a high-pass analog filter?

Solution. In the frequency domain the amplitude loss can (partially) be compensated by means of a first-order high-pass filter, e.g., a voltage divider of two resistors, where the top resistor is shunted with a capacitor. The transfer characteristics are

$$|H_{high-pass}(f)| = \sqrt{1 + 4\pi^2 f^2 \tau^2} = 1 + 2\pi^2 f^2 \tau^2 - \pi^4 f^4 \tau^4 / 2 + \ldots$$

$$|H_{zero-order}(f)| = \frac{\sin(\pi f T_s)}{\pi f T_s} = 1 - \pi^2 f^2 T_s^2/3! + \pi^4 f^4 T_s^4/5! - \ldots$$

With $\tau = T_s/\sqrt{12}$ both functions will compensate for low frequencies, as both second terms add up to zero. However, beyond $f_s/2$ the time-continuous nature of the high-pass filter and the zero-order hold function will no longer match. Moreover the frequency response is an average over infinite time, and the instantaneous time-domain response will show at certain phase relations large excursions. Finally in such a setup the high-frequency noise will be amplified.

Example 2.11. Derive the transfer function for a first-order hold function in Fig. 2.18 (middle).

Solution. The transfer is now:

$$y(t) = x(n-1)T_s + (x(nT_s) - x(n-1)T_s)\frac{t}{T_s}, \qquad nT_s \le t \le (n+1)T_s$$

If the time shift $e^{jn\omega T_s}$ is ignored, the Fourier transform leads to a frequency domain representation for the transfer function:

$$\int_{t=0}^{t=T_s} e^{-j\omega T_s} e^{-j\omega t} dt + \int_{t=0}^{t=T_s} (1 - e^{-j\omega T_s})\frac{t}{T_s} \times e^{-j\omega t} dt = T_s \left(\frac{1 - e^{\omega T_s}}{j\omega T_s}\right)^2 \tag{2.25}$$

Rearranging the terms, extracting the time delay $e^{-j\omega T_s/2}$ and normalizing with T_s yields for the amplitude function:

$$|H(\omega)| = \left(\frac{\sin(\pi f T_s)}{\pi f T_s}\right)^2 \tag{2.26}$$

The first-order reconstruction leads to a better suppression of higher-order aliases.

2.5 Noise

In the previous section the sampling process was analyses from a mathematical perspective. In microelectronics a sampling circuit is realized with a switch and a capacitor. And with that, the standard problems of physical implementation start.

2.5.1 Sampling of Noise

Figure 2.21 and Table 2.2 show an equivalent schematic of the basic sampling circuit consisting of a switch and a storage capacitor. Compared to the ideal situation two non-ideal elements have been added to the switch: the switch resistance R combining all resistive elements between source and capacitor. The resistor is impaired with thermal noise,[12] consequently a noise source is added e_{noise} whose spectrum reaches far[13] beyond the sampling rate of the switch.

$$e_{noise} = \sqrt{4kTRBW} \tag{2.27}$$

with Boltzmann's constant $k = 1.38 \times 10^{-23}$ m^2kgs^{-2} K^{-1} and the absolute temperature T in Kelvin. This formulation expresses the noise in the positive frequency domain from 0 to ∞.

When the switch connects to the capacitor, a low-pass filter is formed by the resistor and the capacitor. The average noise energy on the capacitor is therefore a filtered version of the noise energy supplied by the resistor and is filtered by the complex conjugated transfer function of the RC network. Using standard defined integral tables:

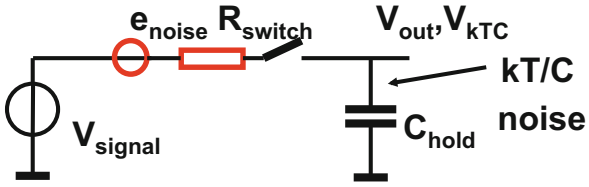


Fig. 2.21 Switched capacitor noise sampling: the series resistances act as a noise source

Table 2.2 kT/C noise for various capacitor values

C_{hold}	$V_{kTC} = \sqrt{kT/C_{hold}}$ at $T = 300\,°C$
10 fF	650 μV_{rms}
100 fF	204 μV_{rms}
1 pF	65 μV_{rms}
3 pF	35 μV_{rms}
10 pF	20.4 μV_{rms}
30 pF	11.8 μV_{rms}

[12] Thermal noise in electronics is modeled as a noise source connected to a real impedance. For circuit calculations this works, but impedances are not noisy, electrons with random energy are.

[13] As thermal noise is an atomic phenomenon, its frequency span ends where sub-atomic mechanisms, as described by quantum physics, start. A rule of thumb limits the standard noise spectrum at 1 THz.

$$v_{C,noise}^2 = \int_{f=0}^{f=\infty} \frac{4kTR\,df}{1+(2\pi f)^2R^2C^2} = \frac{kT}{C} \quad \Rightarrow v_{C,noise} = \sqrt{\frac{kT}{C}} \tag{2.28}$$

The simple and well-known expression for the noise on a capacitor is called[14]: "kT/C-noise." Comparing this result to the power of the sine wave $v_{signal}(t) = \hat{A}\sin(\omega t)$ over the time period $1/\omega$ results in the signal-to-noise ratio SNR:

$$\mathrm{SNR} = \frac{\mathrm{P_{signal}}}{\mathrm{P_{noise}}} = \frac{\widehat{A^2}/2}{kT/C} \tag{2.29}$$

The magnitude of the resistor (the origin of the noise) is not part of this first-order expression. On one hand, an increase of the resistor value will increase the noise energy proportionally, however, that same increase in resistor value will reduce the relevant bandwidth also proportionally.

The same result follows from classical thermodynamics. The equipartition theorem says that in thermal equilibrium, the thermal energy is equally distributed over all degrees of freedom. For a capacitor there is only one degree of freedom: its potential. Therefore energy contained in the thermal fluctuation of carriers $Cv_{C,noise}^2/2$, equals the thermal energy for one degree of freedom: $kT/2$. Solving the equation results again in Eq. 2.28. Obviously there is no resistor involved. In simple terms one can say that the thermo-energetic electrons on the capacitor plates will move every now and then to the voltage source and back again due to their thermal energies. So the voltage over the capacitor fluctuates with time. When the sample switch opens the charge situation freezes.

If the noise spectrum is sampled in Fig. 2.22, each multiple of the sampling frequency will modulate the adjacent noise back to the base band, where all the noise bands accumulate. The same happens to all other bands, thereby hugely increasing the impact of noise.

Equation 2.28 holds for the time-continuous case, where the switch is permanently conductive, but holds equally for the sampled situation. Although the signals look completely different in the time domain, both the time-continuous and the sampled noise signal have values taken from a normal distribution with a zero-value mean and a variance $v_{C,noise}^2 = kT/C$.

This kT/C noise can be interpreted as a flat spectrum in the band from DC to $f_s/2$ as long as the RC cut-off frequency largely exceeds the sample rate. The spectral power noise density (power per Hz) of kT/C noise in a sampled system is equal to kT/C over half of the sample rate:

[14] Very annoying "T" for absolute temperature as well as for fixed time periods.

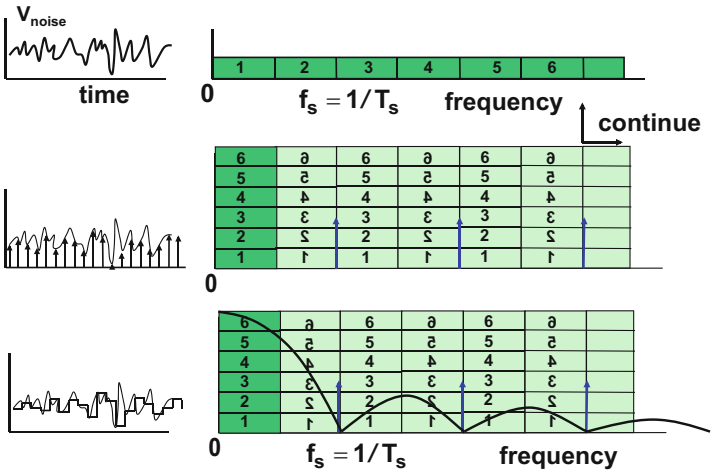


Fig. 2.22 Noise sampling, voltage on the capacitor: *left* in the time domain, *right* in the frequency domain

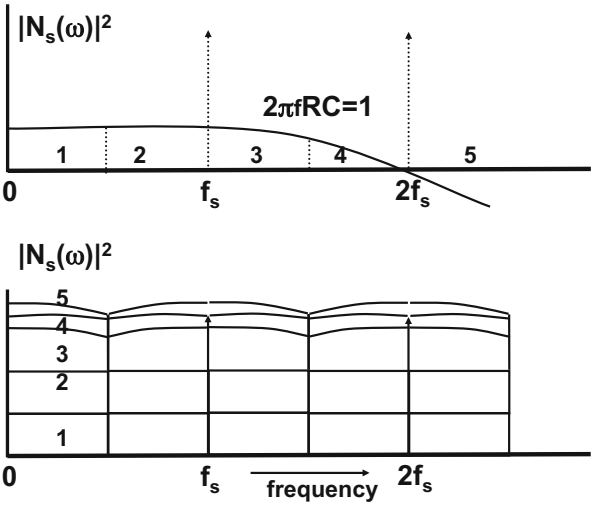


Fig. 2.23 Band-limited noise is sampled in a similar manner as normal signals. *Top*: this noise has a finite bandwidth, *bottom*: after sampling. The power spectra add up and are mirrored

$$S_{ff,SH} = \frac{2kT}{Cf_s} \tag{2.30}$$

If the RC cut-off frequency is low with respect to the sample rate, the noise bandwidth must be treated in a similar fashion as a normal signal band, see, e.g., Fig. 2.23. The power noise density of the time-continuous network with the same resistor and capacitor having a cut-off frequency of $f_{RC} = 1/2\pi RC$ in its pass-band is

$$S_{ff,RC} = 4kTR = \frac{2kT}{\pi C f_{RC}} \tag{2.31}$$

The comparison of the two noise densities in Eqs. 2.30 and 2.31 shows that in the sampling process the noise density increases by a factor $\pi f_{RC}/f_s$. This factor corresponds to the number of bands that stack up in Fig. 2.23. This considerable increase in noise density causes major problems when designing high-resolution converters.

The switching sequence can influence the total noise accumulated in the circuit. In switched capacitor circuits, every switch cycle will add one portion of kT/C noise. As these noise portions mostly are uncorrelated, they will sum in and root-mean-square way. Root-mean-square is the root of the effective power in a sum of signals. Also in situations where a switch discharges the charge of a capacitor into a fixed voltage or even ground potential, kT/C noise will appear (sometimes referred to as "reset-noise").

This kT/C noise term presents a lower boundary in choosing the value for a sampling capacitance. An analog-to-digital converter is signal-to-noise limited because of this choice. A circuit with a total sampling capacitance of 1 pF will be limited by a noise voltage floor of 65 μV_{rms} at room temperature. A larger capacitance value will require IC area, more charging current and will directly impact the power budget.

Example 2.12. An uncorrelated white noise source with a total effective value of 1 mV_{rms} in the band limited to 120 MHz is sampled at 10 Ms/s. What is the noise density of the source? What is the noise density after sampling? What is the rms-value of the noise signal after sampling?

Solution. The effective noise level of 1 mV_{rms} means that the noise has an accumulated power equal to $(1\,mV)^2$ over the impedance. That allows a calculation of the noise density of the noise source: $S_{vv} = (1\,mV)^2/120\,MHz$. After sampling all noise bands higher than $f_s/2$ are folded back to the baseband. In this case the frequency range between DC and 5 MHz will contain 24 uncorrelated noise bands. The noise density is consequently: $S_{vv,s} = 24 \times (1\,mV)^2/120\,MHz$. The total noise after sampling is found from integration of the noise density over the band of 5 MHz, yielding again an effective noise level of: 1 mV_{rms}. What about the noise in the band beyond 5 MHz? The noise density in those bands is equally high and real, but during the reconstruction process no more energy can be retrieved than what is available in one band.

Example 2.13. In a process with a nominal supply voltage of 1.2 V a sinusoidal signal of 100 MHz and 500 $mV_{peak-peak}$ is sampled. A SNR of 72 dB is required. Calculate the sampling capacitor and estimate the circuit power.

Solution. 500 $mV_{peak-peak}$ corresponds to a root-mean-square "rms" voltage of $500/2\sqrt{(2)} = 177\ mV_{rms}$. With a signal-to-noise ratio of $10^{72/20} = 4000$ (corresponding to a 12 bit ADC performance) the kT/C noise must be lower than $177\,mV_{rms}/4000 = 44\,\mu V_{rms}$, and a minimum capacitor of 2.15 pF is needed.

A sinusoidal signal with a frequency of 100 MHz requires a current of $i = \omega \times C \times 500\,\mathrm{mV}_{peak-peak} = 0.675\,\mathrm{mA}_{peak-peak}$. This charge on the capacitor has to be supplied from an electronic circuit that allows only current in one direction, a bias current of, e.g., 1 mA can be used. Now the current swings from 162.5 to 837.5 μA. In first order this circuit requirement will consume 1.2 mW.

Example 2.14. A signal is sampled on two parallel connected equal capacitors: C_1, C_2 and $C_2 = C_1$. After sampling the capacitors are stacked in order to double the signal voltage, see Fig. 2.24. Does the signal-to-noise ratio change between the parallel and stacked connection?

Solution. After sampling a voltage V is stored on each capacitor. A noise contribution $v_{noise} = kT/(C_1 + C_2)$ is added and the signal-to-noise ratio is determined by the rms value of the signal over the noise. This noise gets "frozen" after the sample switch is opened, the same value holds for both capacitors, this is a rare situation where the noise is correlated. In first approximation the stacked capacitor construction will double the signal and its rms value, but the noise contribution on both capacitors doubles too, as these are correlated. The signal-to-noise ratio remains the same.

Ready? Not so fast, time for a second look!

There is another sampling moment that arises when the connection between both capacitors is opened in order to perform the stacking. Until that moment, the electrons with their thermodynamic energy can move freely between the capacitors. When the connection between the capacitors is broken, a new kT/C sampling event happens. This time the noise between the top plates will be equivalent to the noise of the series connection of C_1 and C_2, which equals $2kT/C_1$ if the capacitors are equal. The voltage over every capacitor is kT/C_1 and both voltages are correlated, but with opposite sign. After stacking these two opposing noise contributions will cancel! And the signal-to-noise ratio will indeed be the same.

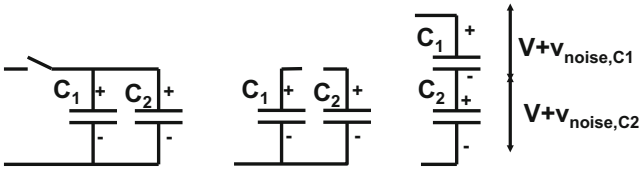


Fig. 2.24 A signal is sampled on two equal capacitors, after which the capacitors are stacked to double the signal

2.6 Jitter

2.6.1 Jitter of the Sampling Pulse

In the previous analysis it is assumed that the sample moments are defined with infinite precision. In practice all signals that define time moments have limited bandwidths, which means that there is no infinitely sharp rising edge. Oscillators, buffers, and amplifiers are all noisy devices [23–26], so consequently they add noise to these edges in Fig. 2.25. If noise changes the switching level of a buffer, the outgoing edge will have a varying delay with respect to the incoming edge. This effect is called: jitter. Jitter causes sample moments to shift from their position, and consequently the sampling circuit will sample the signal at another time moment. Next to noise-like components also signal-related components may influence the clock edge through limited power supply rejection, capacitive coupling, etc. Jitter from noisy sources will result in noise contributions to the signal, jitter from deterministic sources leads to tones (from fixed carriers) or to distortion (if the jitter source is correlated to the signal). Examples of systematic offsets in timing are: skews due to unequal propagation paths of clocks, interference from clock dividers, and clock doubling by means of edge detection. Random "jitter" variations occur not only during the generation of clock signals in noise-sensitive oscillators and PLLs, but also during transportation of timing signals jitter can be added, e.g., in long chains of clock buffers fed by noisy digital power supplies, capacitive coupling, and varying loading. A practical value for jitter on a clock edge coming from a digital CMOS environment[15] is 30–100 ps_{rms}. If an advanced generator is used in combination with bandpass filters, the sampling pulse jitter can be reduced to levels below 100 fs_{rms}. The contributions of dedicated high-power on-chip circuits can be

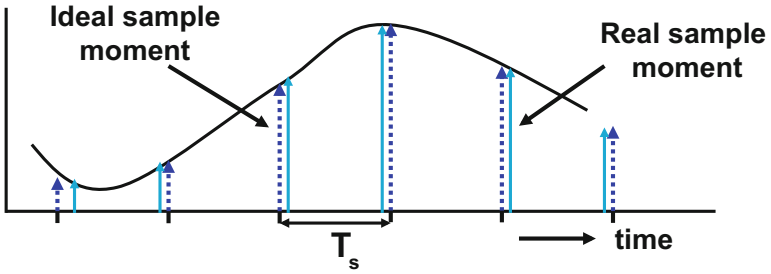


Fig. 2.25 The ideal sampling moments (*dashed*) shift in an arbitrary fashion in time if the sample clock is disturbed by jitter

[15] A peak–peak value is often used for jitter, but peak–peak values for stochastic processes have no significance if the process and the corresponding number of observations are not identified.

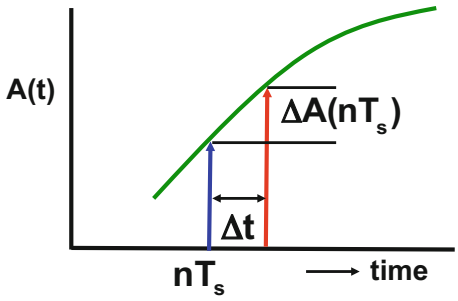


Fig. 2.26 The ideal sampling moment is affected by jitter and an amplitude error occurs

brought back to a similar level. For example, Ali et al. [219] reports an overall jitter of 83 fs_{rms}.

The above description specifies the cycle-to-cycle deviation. In some systems a long-term jitter component can be relevant, e.g., for the display of a signal on a screen via a scanning mechanism, the jitter between two samples in the scan direction is determined by cycle-to-cycle jitter, while two samples arranged above each other are given by a long-term jitter. In monitors these samples can be some 1000 clock cycles apart. This jitter is specified over a longer period and requires some extensions of the following analysis. Some more in the discussion of phase-noise in Sect. 2.6.2.

Figure 2.26 shows the effect of shifting a sample moment. If a sinusoidal signal $A(t) = \hat{A}\sin(\omega t)$ with a radial frequency ω is sampled by a sample pulse with jitter, the new amplitude and the amplitude error are estimated as:

$$A(nT_s + \Delta t(t)) = \hat{A}\sin(\omega \times (nT_s + \Delta t(t)) \tag{2.32}$$

$$\Delta A(nT_s) = \frac{d\hat{A}\sin(\omega t)}{dt} \times \Delta t(nT_s) = \omega\hat{A}\cos(\omega nT_s)\Delta t(nT_s) \tag{2.33}$$

The time error is a function of the time itself. The amplitude error is proportional to the slope of the signal $\omega\hat{A}$ and the magnitude of the time error.

If the time error is replaced by the standard deviation σ_t^2 describing the timing jitter variance, the standard deviation of the amplitude σ_A is estimated as:

$$\sigma_A^2(nT_s) = \left(\frac{dA(nT_s)}{dt}\right)^2 \sigma_t^2 = \omega^2\hat{A}^2\cos^2(\omega nT_s)\sigma_t^2 \tag{2.34}$$

Averaging this result over all values of nT_s gives a jitter error power of:

$$\sigma_A^2 = \frac{\omega^2\hat{A}^2\sigma_t^2}{2} \tag{2.35}$$

When the origin of the jitter is a flat spectrum as for thermal noise, this jitter noise will appear as a flat spectrum between 0 and $f_s/2$ and repeats at every higher band.

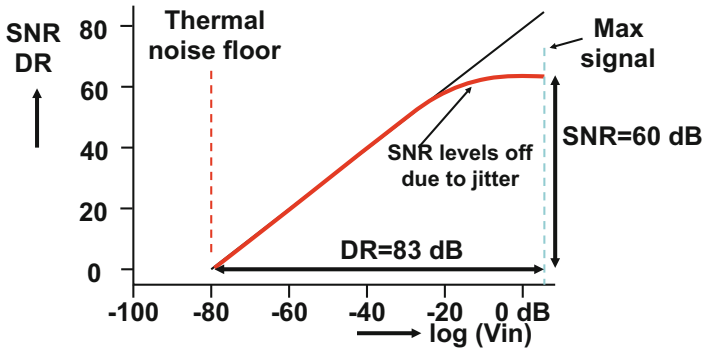


Fig. 2.27 With increasing signal amplitude over a fixed thermal noise level the signal-to-noise ratio increases. However when the amplitude is such that jitter becomes the dominant error, the signal-to-noise ratio flattens

Comparing this result to the power value of the sine wave $\widehat{A^2}/2$ over the time period $T = 1/\omega$ results in the signal-to-noise ratio due to jitter:

$$\text{SNR} = \frac{\text{P}_{\text{signal}}}{\text{P}_{\text{jitter}}} = \frac{\widehat{A^2}/2}{\sigma_A^2} = \left(\frac{1}{\omega\sigma_t}\right)^2 = \left(\frac{1}{2\pi f\sigma_t}\right)^2 \tag{2.36}$$

or in deciBel[16] (dB):

$$\text{SNR} = 20\,{}^{10}\log\left(\frac{1}{\omega\sigma_t}\right) = 20\,{}^{10}\log\left(\frac{1}{2\pi f\sigma_t}\right) \tag{2.37}$$

For sampled signals the above relations hold for the ratio between the signal power and the noise in half of the sampling band: $0\ldots f_s/2$. This simple relation estimates the effect of jitter, assuming no signal dependencies. Nevertheless it is a useful formula to make a first order estimate. For wide-band signals with a uniform power distribution between $0,\ldots,f$ [27] gives a 3× higher signal-power to noise ratio or a 4.8 dB more favorable jitter SNR. Note that the jitter power is independent of the sample rate, consequently the jitter power density (power per Hertz) is inversely related to f_s

The linear dependence of jitter noise to the input frequency and to the signal amplitude often allows a rapid identification of jitter in a time-discrete system. For a given signal frequency the jitter power increases linearly with the amplitude, leading to the flattening of the SNR versus input amplitude curve [28], see Fig. 2.27.

[16]For some reason deciBel is spelled with single "l" although it was named after A.G. Bell. Similarly the letter "a" was lost in "Volta."

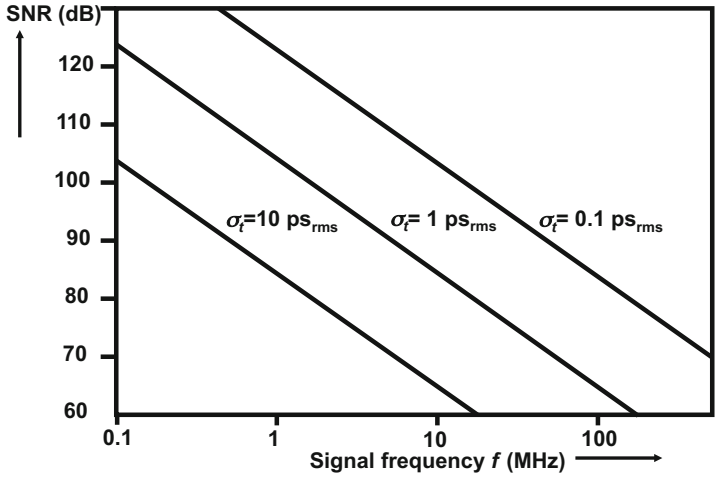


Fig. 2.28 The signal-to-noise ratio depends on the jitter of the sampling signal and the frequency of the time-continuous signal

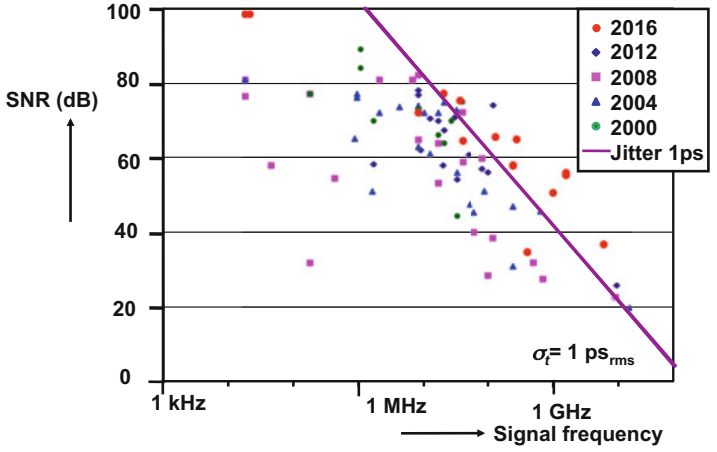


Fig. 2.29 The signal-to-noise ratio versus the signal frequency of analog-to-digital converters reported on the International Solid-State Circuits conferences in the years 2000, 2004, 2008, 2012, and 2016. (from: B. Murmann, "ADC Performance Survey 1997–2016," Online: http://web.stanford.edu/~murmann/adcsurvey.html)

Figure 2.28 shows the signal-to-noise ratio as a function of the input frequency for three values of the standard deviation of the time jitter.

Figure 2.29 compares the jitter performance of analog-to-digital converters published on the International Solid-State Circuits conferences in the years 2000, 2004, 2008, and 2012. It is obvious that a jitter specification better than $\sigma_t < 1$ ps

Table 2.3 Jitter specifications of some commercially available parts

Part	Description	Jitter
"2011"	Quartz 50–170 MHz	3 ps_{rms}
"8002"	Programmable oscillator	25 ps_{rms}
"1028"	MEMS+PLL combi 100 MHz	95 ps_{rms}
"6909"	RC oscillator 20 MHz	0.2 %
"555"	RC oscillator/timer	>50 ns_{rms}

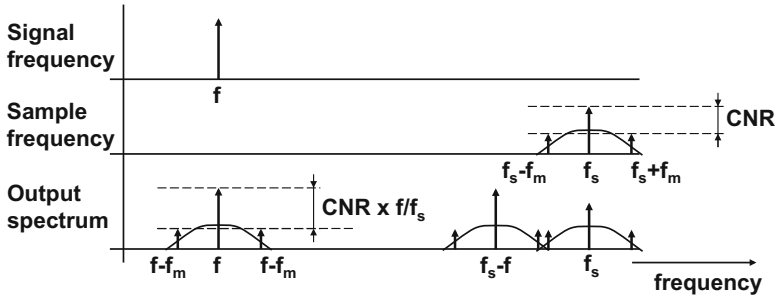


Fig. 2.30 Jitter around the sampling frequency will produce side spectra around the input tone

is a challenge. The comparison of the best converters in every year shows that little progress was made over the last decade.

Table 2.3 indicates some jitter numbers from commercial timing components.

From a spectral point of view, the jitter spectrum modulates the input tone. Therefore the jitter spectrum around the sampling pulse will return around the input frequency as in Fig. 2.30. Translated to a lower frequency the time error due to jitter will produce a proportionally smaller amplitude error. Therefore the carrier-to-noise ratio (CNR) improves.

Equation 2.35 has been derived for a single sine wave as a signal. In communication systems (from ADSL to 5G) multi-tone signals are applied. These signals contain large number of carriers N_c up to 1024. All carriers behave as independent sine waves, and the probability that all carriers are at a maximum is far below the system error level of $10^{-4} - 10^{-6}$, as is typical for these systems. Instead of allowing an individual carrier amplitude of just $\hat{A}/N_c$, more commonly the amplitude of the i-th carrier is approximated by Zogakis and Cioffi [29] and Clara and Da Dalt [30]

$$\hat{A}_{c,i} = \frac{\hat{A}}{C_F \sqrt{N_c}}$$

C_F is the so-called crest-factor: the ratio between the maximum signal and the rms value. In ADSL $C_F \approx 5.6$. The jitter error power per carrier is given by Eq. 2.35. Summing the power over all N_c carriers yields

$$\sum_{i=1}^{i=Nc} \sigma_{Ac,i}^2 = \sum_{i=1}^{i=Nc} \frac{\omega_i^2 \hat{A}^2 \sigma_t^2}{C_F^2 N_c} \approx \frac{\omega_{middle}^2 \hat{A}^2 \sigma_t^2}{C_F^2} \tag{2.38}$$

In this coarse approximation ω_{middle} is the frequency of the tone at $i = N_c/2$. Obviously the jitter error power is far lower than in the sine wave case. More accurate analysis is found in [29, 30].

If jitter is caused by delay variations in digital cells as shown in Fig. 2.31, the jitter can also contain signal components and strong spurious components, e.g., linked to periodic processes in the digital domain. These contributions are demodulated similar as in Fig. 2.30 and are the source for spurious components and signal distortion. Therefore digital circuits that generate and propagate the sample pulse must be treated as if these were analog blocks.

Example 2.15. The clock buffer in Fig. 2.31 has edge transition times of 70 ps. How much jitter can be expected if a random noise of 60 mV_{rms} is present on the power supply of 1.2 V.

Solution. Due to voltage changes on the power supply lines, the currents inside the buffer will change, in first order proportional to the voltage change. As a consequence the slope of the transition will vary linearly. The mid-level switching point is now reached after 35 ps delay from the input mid-level passing. A voltage change of 60 mV/1.2 V= 5 % will create a 5 % delay variation on the slope and the delay of 35 ps. So the expected jitter is 1.75 ps_{rms} per edge. As the same voltage variation applies to two inverters, the overall jitter is 3.5 ps_{rms}.

Example 2.16. In Example 2.6 the sample sequence is distorted by a random jitter component of 5 ps_{rms}. Is it possible to discriminate in the sampled data domain between a 15 MHz input sine wave or a 135 MHz input sine wave?

Solution. With perfect sampling both signals will result in equivalent wave forms in the sampled data domain. However, the presence of jitter allows to discriminate,

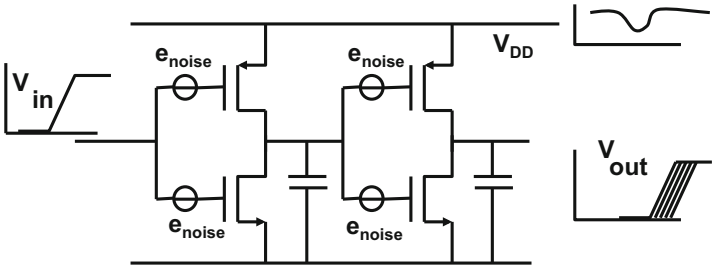


Fig. 2.31 A clock buffer for generating the digital sample signal can add to an ideal sample signal some noise of the buffer transistors. Also fluctuations on the power supply will affect the switching behavior of the buffer, causing uncertainty on the edges and jitter in the sampling

as the resulting SNR for a 15 MHz input signal is $SNR = 1/2\pi f_i \sigma_t = 66.5\,\text{dB}$, while the SNR for 135 MHz equals 47.5 dB.

Example 2.17. Calculate the jitter due to thermal noise that an inverter with dimensions of NMOST 0.2/0.1 and PMOST 0.5/0.1 in a 90-nm CMOS process adds to an edge of 50 ps (bottom-top).

Solution. Every transistor adds noise that is related to the transconductance in the channel: $i_{noise} = \sqrt{4kTBWg_m}$. If the inverter is at its mid-level point (0.6 V) both transistors will be contributing to a total noise current of: $i_{noise,n+p} = \sqrt{4kTBW(g_{m,n} + g_{m,p})}$. This noise corresponds to an input referred noise voltage of $v_{noise,n+p} = i_{noise,n+p}/(g_{m,n} + g_{m,p}) = \sqrt{4kTBW/(g_{m,n} + g_{m,p})}$. With the help of the parameters in Table 4, an equivalent input noise voltage is found of 0.62 mV_{rms} in a 10 GHz bandwidth. This bandwidth is an approximation based on the observation that an rising edge of 50 ps followed by a similar falling edge limits the maximum frequency of the inverter to 10 GHz. The jitter order of magnitude is estimated as: $\sigma_t/\tau_{edge} = v_{noise,n+p}/V_{DD}$ and $\sigma_t = 25\ \text{fs}_{rms}$.

2.6.2 Phase-Noise and Jitter

For sampling systems the variation in time moments or jitter is an important parameter. Jitter is here described as a random time phenomena. In RF systems the same phenomenon is observed in the frequency domain and is called "phase-noise." The events in oscillator and PLL spectra, such as in Fig. 2.32, are specified at the offset frequency with respect to the ideal oscillation frequency f_o. A spectrum of the signal from a phase-locked loop or oscillator circuit shows some typical components, see Fig. 2.32:

- White noise in the output (no dependency on the frequency).
- White noise that modulates the oscillator shows up in the power spectrum with a decreasing frequency slope in the oscillation offset frequency. $1/f$ noise generates an even faster decreasing slope in the spectrum with a low offset frequency.

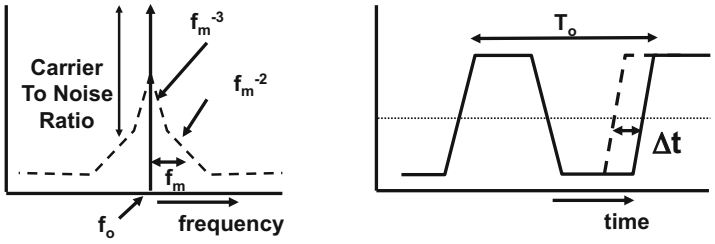


Fig. 2.32 *Left*: the frequency spectrum of an oscillator, *right*: time jitter

- PLLs multiply a reference frequency. Often spurious tones are visible on both sides of the generated frequency at an offset frequency equal to the reference frequency.
- Undesired tones entering the PLL via substrate coupling can modulate the output.

An instantaneous phase deviation $\Delta\theta$ offsets the zero-crossings of a sinusoidal signal of frequency ω_o to yield a time deviation. The time variation Δt for a radial frequency ω_o with a corresponding frequency period T_o, and the phase deviation $\Delta\theta = \omega_o \Delta t$ of the same signal, essentially describe the physical phenomena [28, 31]:

$$\frac{\Delta t}{T_o} = \frac{\Delta\theta}{2\pi} \tag{2.39}$$

From this instantaneous relation between time offset and phase offset, a first-order indication of the relation between jitter and phase-noise is obtained. Both originate from the same stochastic source. Now the time-domain offset Δt is replaced by its time-averaged variance: $\sigma^2_{t,rms}$. In order to obtain the phase-error variance $\sigma^2_{\theta,rms}$, the spectral noise density $S_{ff}(f)$ must be integrated over both side lobes to give the total equivalent phase noise power[17]:

$$\left(\frac{\sigma_{t,rms}}{T_o}\right)^2 = \left(\frac{\sigma_{\theta,rms}}{2\pi}\right)^2 = \frac{2\int_{f_{low}}^{f_{high}} S_{ff}(f)df}{(2\pi)^2} \tag{2.40}$$

The span of integration is limited by a lower and higher boundary of the offset frequency. The integration cannot start at $\omega = \omega_o$ Hz due to the singularity in the spectral density. Leaving out the frequencies below 10^{-8} means ignoring 3-years repetitive effects, but more often a lower boundary is chosen in the Hz to kHz range. Obviously one should not expect a 99.9 % prediction level. The choice for f_{low} also depends on whether the cycle-to-cycle jitter is required or longer-term jitter variations. The energy in the low-frequency second-order lobes of the phase spectrum is responsible for the increase of long-term jitter over cycle-to-cycle jitter. In the extreme case of only white phase noise, the contribution of the low-frequency band would be negligible and the long-term jitter would be comparable to the cycle-to-cycle jitter. Translating various forms of phase-noise densities in time jitter clearly requires an assumption of spectral density function for the phase noise [23–26, 31].

Example 2.18. Calculate the jitter from the spectrum in Fig. 2.33.

[17] Here a strictly formal derivation requires some 10 pages, please check out specialized literature, e.g., [31].

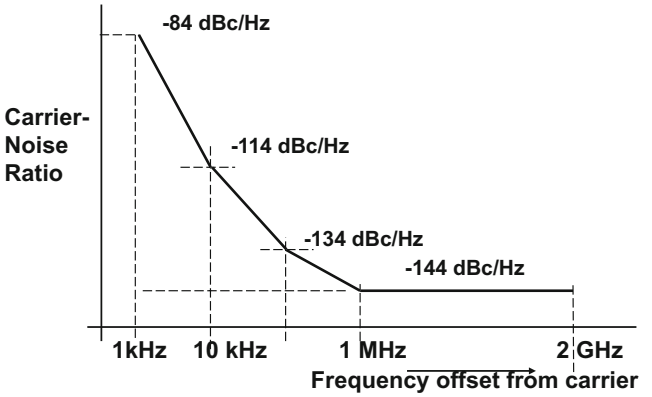


Fig. 2.33 The phase-noise spectrum of an oscillator signal at 2 GHz

Solution. f is the offset frequency from the carrier. The spectrum shows four typical regions: from $f = 1\,\text{kHz}$ to 10 kHz the slope is f^{-3} corresponding to the 1/f noise in the oscillator, the next section with a slope of f^{-2} is due to thermal noise in the oscillator. The two right most regions with slope f^{-1} and the floor are due to 1/f noise and thermal noise in buffers. This spectrum appears on either side of oscillator frequency f_0. The level of the floor is given as $-144\,\text{dBc/Hz}$. Corresponding to a power of 4.10^{-15} of the carrier per Hz bandwidth. The curve is approximated with the following equation:

$$S_{ff}(f) = \left[1 + \frac{f_1}{f} + \left(\frac{f_2}{f}\right)^2 + \left(\frac{f_3}{f}\right)^3\right] S_{floor}(f) \tag{2.41}$$

where $f_1 = 1\,\text{MHz}$, $f_2 = 100\,\text{kHz}$, $f_3 = 10\,\text{kHz}$ and the range of interest is limited from 1 kHz to 2 GHz. Formally the phase area under the curve is found by integration and substitution of the frequencies.

$$\left(f + f_1 \ln(f) + f_2^2/f + f_3^3/2f^2\right)\ S_{floor}\Big|_{f=1\,kHz}^{2\,GHz} \tag{2.42}$$

A coarse approximation allows to determine the contribution in each section of the curve, which gives an insight where optimization of the circuit is most beneficial:

$$\begin{aligned}
&\left((2.10^9 - 10^6) + f_1(\ln(10^6) - \ln(10^5)) + f_2^2(1/10^4 - 1/10^5)\right.\\
&\left. + f_3^3(1/10^3 - 1/10^4)/2\right) 4.10^{-15}\\
&= \left(1999 \times 10^6 + 2.3 \times 10^6 + 0.9 \times 10^6 + 450 \times 10^6\right) 10^{-15}\\
&= 9.8 \times 10^{-6}
\end{aligned}$$

With the help of Eq. 2.40 the time jitter is found: 0.35 ps_{rms}. This is a real coarse estimate, e.g., a popular spread sheet called the "Allen Variance" simply adds 3 dB to the noise density to compensate for underestimations. Note that the thermal noise and the 1/f noise in the oscillator dominate.

2.6.3 Optical Sampling

The $\sigma_t = 0.1, \ldots, 10$ ps_{rms} range for jitter is typical for electronic design and ultimately linked to physical processes such as thermal noise and $1/f$ noise. Mode-locked lasers can generate pulse trains with 200 ps width and a jitter of approximately ten femtoseconds. Building sampling devices triggered by these lasers is a challenge, as the straight-forward solution to capture the laser pulses with diodes would immediately affect the performance. An alternative solution [32] uses GaAs finger structures and attributes $\sigma_t = 80$ fs_{rms} jitter to the sampling process.

2.7 Time-Discrete Filtering

Time-discrete filtering forms a subset of the time-discrete signal processing tool box, see, e.g., [16, 17] and can be found in oversampled digital-to-analog converters, Sect. 10.1, and in sigma-delta modulators, Sect. 10.4. Time-discrete filters play a role in the conversion architecture decisions as well as in the necessary post-processing.

2.7.1 FIR Filters

Sampled signals can easily be delayed in the time-discrete domain. In the analog time-discrete domain, switched capacitors transfer charge packets from one stage into another stage. By means of appropriate switching sequences various time delays are implemented. After amplitude quantization samples can also be delayed in the digital domain via digital delay cells, registers, and memories. Frequency filters in each domain are realized by combining the time-delayed samples with specific weighting factors or multiplying coefficients.

Operations and functions in the discrete-time domain are described in the z-domain. If $f(n) = f(nT_s), n = 0 \ldots \infty$ is a sequence of values corresponding to the sample value of $f(t)$ at points in time $t = nT_s$, this sequence can be described in the z-domain as:

$$f(z) = \sum_{n=0}^{n=\infty} f(n) z^{-n}$$

where z is a complex number in polar representation $z = re^{j\omega_z}$, which resembles the Laplace parameter $s = \alpha + j\omega$ with $r \leftrightarrow e^{\alpha}$, $\omega \leftrightarrow \omega_z$. The important difference is that the z-domain describes a sampled system where the maximum frequency is limited to half of the sample rate. While ω is expressed in rad/sec, $\omega_z \leftrightarrow \omega T_s$ is expressed in radians and abstracts from physical frequencies. The s-plane and the z-plane can be mapped on each other. Due to the polar description the $j\omega$ axis in the s-domain becomes a unity circle in the z-domain, with the DC point at $z = 1e^{j0} = 1$. Poles and zeros in left-side of the s-plane resulting in stable decaying exponential functions in the time domain move to the inner part of the unity circle in the z-domain, Fig. 2.34 (right).

A delay of one basic sample period is transformed into the function z^{-1}. A frequency sweep from 0 to $f_s/2$ results in a circular movement of the z vector in a complex plane from $+1$, via $0 + j$ to -1. For the frequency range $f_s/2$ to f_s the z vector will turn via the negative imaginary plane and return to $z = 1$. Figure 2.35 shows two integrators described in the z domain. The left structure adds the present sample to the sum of the previous samples. After the next clock the output will equal that sum and a new addition is performed. The right topology does the same, here the sum is directly available. The transfer functions for both structures are

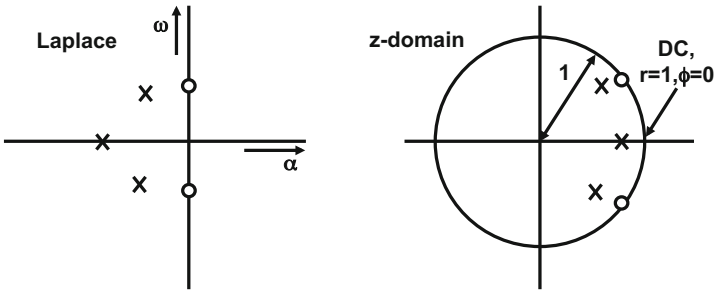


Fig. 2.34 The complex plane for the Laplace transform (s-plane) and the time-discrete plane (z-plane). A real pole, a pair of imaginary poles and a pair imaginary zeros are depicted

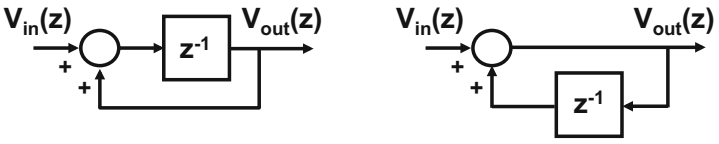


Fig. 2.35 Two integrators in the z-domain

$$H(z) = \frac{z^{-1}}{1 - z^{-1}} \qquad H(z) = \frac{1}{1 - z^{-1}}$$

These integrator formulas indicate a mathematical pole at $z = 1$. The transform to the Laplace domain $z \leftrightarrow e^{sT_s}$ shows that $z = 1$ corresponds to $s = 0$ or DC conditions. And indeed a DC signal on an ideal integrator will lead to an unbounded output. Close to $z = 0$ the left integrator has zero output while in right-hand integrator just passes the signal.

The most simple filter in the time-discrete domain is the comb filter. The addition of a time-discrete signal to an m-sample periods delayed signal gives a frequency transfer that can be evaluated in the z-domain:

$$H(z) = 1 \pm z^{-m}$$

This function has zeros at all frequencies where $z^{-m} = \pm 1$, resulting in m zeros distributed over the unity circle. Using the approximation $z \leftrightarrow e^{sT_s}$ results in:

$$\begin{aligned} H(s) &= 1 \pm e^{-smT_s} = e^{-smT_s/2}(e^{+smT_s/2} \pm e^{-smT_s/2}) \\ |H(\omega)| &= 2|\cos(\omega mT_s/2)|, \quad \text{addition} \\ |H(\omega)| &= 2|\sin(\omega mT_s/2)|, \quad \text{subtraction} \end{aligned} \tag{2.43}$$

where the sign at the summation point determines whether the cosine response (with equal signs) or the sine response (with opposite signs) applies, see Fig. 2.36. In this plot the zeros are observed in the frequency domain.

Comb filters are mostly applied in systems where interleaved signals have to be separated. An example is the analog composite video signal, where the frequency carriers with the color information are interleaved between the carriers for the luminance signal.

The comb filter adds signals to their delayed versions. A more general approach uses a delay line where each delayed copy is multiplied with its own weight factor,

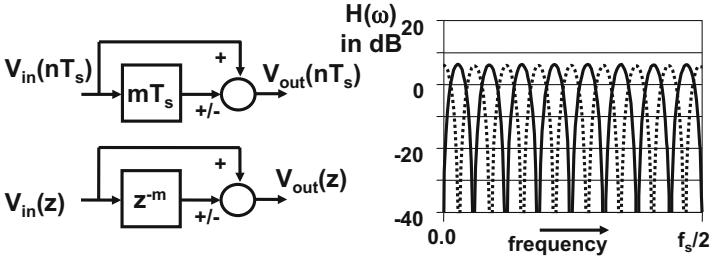


Fig. 2.36 The comb filter as sampled data structure and in the z-domain. The frequency response shows with a *solid line* the sine response (minus-sign at the summation), while the *dotted line* represents the cosine response (plus-sign)

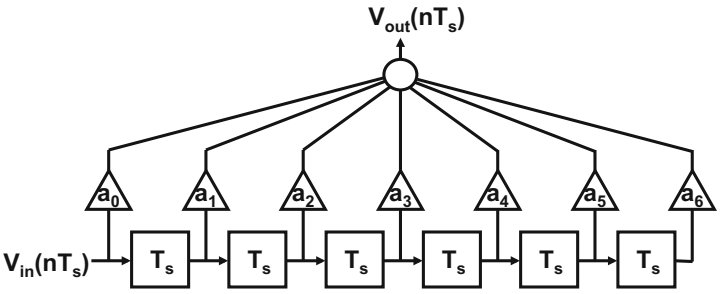


Fig. 2.37 The general structure of a finite impulse filter

see Fig. 2.37. A filter with this structure is known as a "Finite Impulse Response" filter (FIR-filter). The term "finite" means that any input disappears from the filter after passing through the N delay elements. In the summation the signals from the different delay elements can enhance or extinguish each other depending on the periodicity of the signal with respect to the delay time and the multiplication factor. The filter designer can adapt the filter characteristic through these multiplication coefficients or weight factors. Similar to the time-continuous filters the discrete-time transfer function defines the relation between input, filter transfer function, and output in the time domain with a convolution:

$$y(nT_s) = \sum_{k=0}^{k=\infty} h(k)x(nT_s - kT_s) \tag{2.44}$$

Applied to the filter in Fig. 2.37, this gives:

$$V_{out}(nT_s) = \sum_{k=0}^{k=N-1} a_k V_{in}((n-k)T_s) \tag{2.45}$$

An intuitive way of realizing what happens in an FIR filter is to imagine an endless row of delayed samples. Over this row a window defined by the FIR filter length is moved.[18] The z-transform results in a description of the transfer of an FIR filter:

$$\frac{V_{out}(z)}{V_{in}(z)} = H(z) = \sum_{k=0}^{k=N-1} a_k z^{-k} \tag{2.46}$$

In order to transform this transfer function from the discrete time domain to the frequency domain, the term z^{-1} is substituted by $e^{-j\omega T_s}$ which results in:

[18] In financing a monthly moving average is a very simple FIR filter with 12 taps and simply "1" as multiplication factor.

$$H(\omega) = \sum_{k=0}^{k=N-1} a_k e^{-jk\omega T_s} \tag{2.47}$$

This time-continuous approximation is only applicable for a frequency range much smaller than half of the sample rate.

Some important properties of this filter are related to the choice of the weighting factors. Suppose the values of the coefficients are chosen symmetrical with respect to the middle coefficient. Each symmetrical pair will add delayed components with an average delay equal to the middle position. If the delay of each pair equals $NT_s/2$, then the total filter delay will also equal $NT_s/2$. The same arguments holds if the coefficients are not of equal magnitude but have an opposite sign ("anti-symmetrical"). This "linear phase" property results in an equal delay for all (amplified or attenuated) signal components and is relevant if the time-shape of the signal must be maintained, e.g., in quality audio processing.[19]

Mathematically the constant delay or linear phase property can be derived from Eq. 2.47 by substitution of the Euler's relation[20]:

$$e^{-j\omega T_s} = \cos(\omega T_s) - j\sin(\omega T_s) \tag{2.48}$$

After moving the average delay $NT_s/2$ out of the summation, real and imaginary terms remain:

$$H(\omega) = e^{-j\omega NT_s/2} \sum_{k=0}^{k=N/2-1} (a_k + a_{N-k})\cos(k\omega T_s/2) - j(a_k - a_{N-k})\sin(k\omega T_s/2) \tag{2.49}$$

Without violating the general idea, N has been assumed here to be even. If the coefficients a_k and a_{N-k} are equal as in the symmetrical filter the sine term disappears. The cosine term is removed by having opposite coefficients in an asymmetrical filter. Both filters have a constant delay. Depending on the symmetry and the odd or even number of coefficients the filters have structural properties, e.g., an asymmetrical filter with an even number of coefficients has a zero DC-transfer.

A filter that averages over N samples is designed with coefficients of value $1/N$. A simple transfer characteristic can be determined by hand for a small number of coefficients. More complex filters require an optimization routine. A well-known routine was proposed by McClellan, Parks, and Rabiner (MPR or the "Remez exchange algorithm") [33]. This routine optimizes the transfer based on a number of filter requirements.

[19] The human ear is sensitive to delay variations down to the microsecond range.

[20] The definition for Euler's relation is: $e^{j\pi} + 1 = 0$. According to Feynman this is the most beautiful mathematical formula as it is relates the most important mathematical constants to each other.

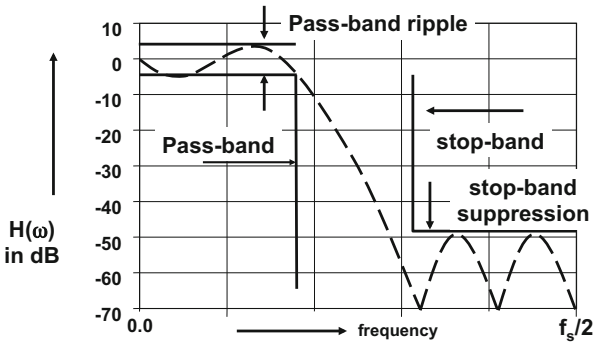


Fig. 2.38 Definition scheme of a filter response

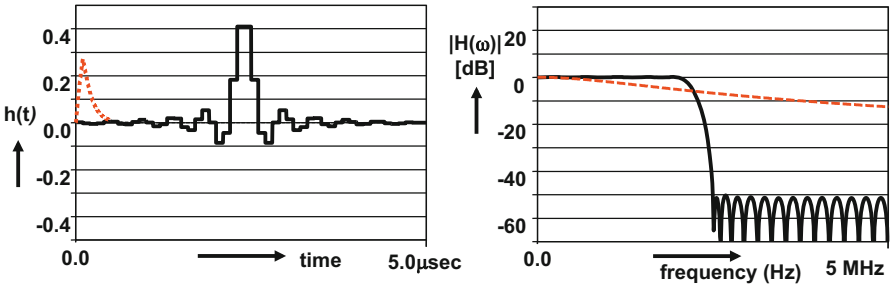


Fig. 2.39 A low-pass FIR filter with 48 coefficients, *left* is the impulse response of the filter and its analog realization (*dashed*). *Right* is the frequency response of both plotted on a linear frequency scale

Figure 2.38 shows a number of terms to define various specification points. Next to that the number of delay elements N, the number of pass and stop bands, and the form of the transition between the bands are required. Some variants of filter design programs allow to compensate alias filters or zero-order hold-effects.

Figure 2.39 shows the impulse response for a somewhat more elaborate filter with 48 coefficients. An impulse response is obtained by shifting a "1" value through the structure preceded and followed by zero samples. An impulse response reveals all filter coefficients. A 2-pole RLC filter transfer function with a quality factor of 0.5 is drawn in dotted lines as a comparison. The delay time is of course much shorter than the 24 cycles of the FIR filter. However the suppression of the digital filter is superior to a simple analog filter[21] or a 7-tap filter as in Fig. 2.41.

Redesigning this filter with the same 10 Ms/s sample rate and 48 coefficients creates a bandpass filter, Fig. 2.40.

The digital time response highly resembles the ringing of a high-Q analog filter of the same specification. The accuracy in which required filter characteristics can be

[21] An equivalent analog filter would require 10–12 poles.

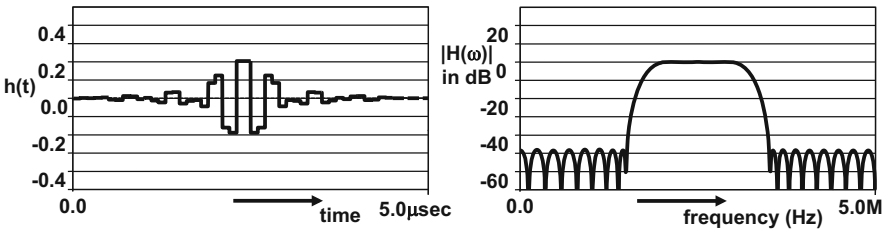


Fig. 2.40 A bandpass FIR filter with 48 coefficients, *left* is the impulse response of the filter and its analog realization. *Right* is the frequency response of both

Table 2.4 Coefficient values for the low-pass FIR filter of Fig. 2.41

Coefficient	Value
$a_0 = a_6$	−0.06
$a_1 = a_5$	0.076
$a_2 = a_4$	0.36
a_3	0.52

defined with FIR filters is clearly illustrated here. In practical realizations the price for an accurately defined filter is the large hardware cost of the delay elements, the coefficients and their multipliers, and the associated power consumption.

The FIR filter has been described in this section as a mathematical construction and no relation was made with the physical reality. Some examples of fully analog FIR realizations are found in switched capacitor circuits and charge-coupled devices.[22] Most FIR filters are implemented in the digital domain: from IC building blocks to FPGA and software modules. In digital-to-analog conversion the semi-digital filter uses digital delays with analog coefficients, see Sect. 7.3.7.

Example 2.19. Determine with a suitable software tool the coefficients for the structure in Fig. 2.37 to create a low-pass filter.

Solution. If the transition for the low-pass filter is chosen at approximately $f_s/4$ coefficients as in Table 2.4 is found.

Figure 2.41 shows the time response and the frequency transfer function from Fig. 2.37 with the coefficients of Table 2.4. In this example of a time-discrete filter the frequency transfer is symmetrical with respect to half of the sampling rate, which was chosen at 10 Ms/s. The spectrum repeats of course at multiples of the sampling rate.

[22] In the period 1970–1980 the charge-coupled device was seen as a promising candidate for storage, image sensing, and signal processing. Analog charge packets are in this multi-gate structure shifted, split and joint along the surface of the semiconductor. Elegant, but not robust enough to survive the digital era.

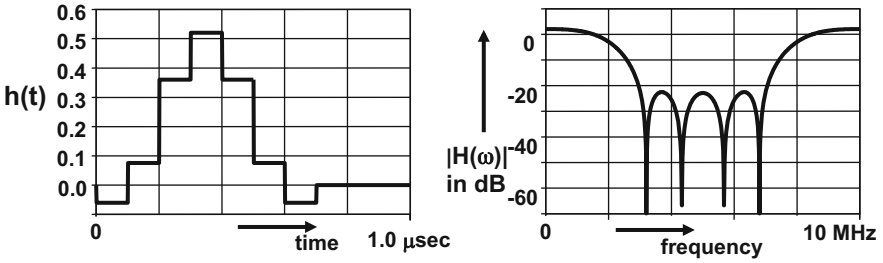


Fig. 2.41 The impulse response and the frequency transfer function of a seven coefficient filter from Fig. 2.37 at a 10 Ms/s sample rate

2.7.2 Half-Band Filters

In order to reduce the complexity of digital FIR filters additional constraints are needed. Introducing the symmetry requirement:

$$H(\omega) + H(\omega_s/2 - \omega) = 1 \tag{2.50}$$

leads to such a complexity reduction. At a frequency $\omega = \omega_s/4$ this constraint results in $H(\omega_s/4) = 0.5$, while the simplest fulfillment of the symmetry requirement around $\omega_s/4$ forces a pass-band, on one side, and a stop band, on the other side, of this quarter sample rate. Consequently these filters are known as "half-band" filters. Substitution of the transfer function for symmetrical filters with an odd number of N coefficients $k = 0, 1, \ldots, m, \ldots, N-1$ and with the index of the middle coefficient equal to $m = (N-1)/2$, leads to:

$$\begin{aligned} a_m &= 0.5 \\ a_{m+i} = a_{m-i} &= C_i, \qquad i = 1, 3, 5, \ldots \\ a_{m+i} = a_{m-i} &= 0, \qquad i = 2, 4, 6, \ldots \end{aligned}$$

Half of the filter coefficients are zero and need no hardware to implement. Optimizing the filter transfer for a minimum deviation of an ideal filter results in a $\sin(x)/x$ approximation:

$$a_{m+i} = \frac{\sin(i\pi/2)}{i\pi}, \qquad i = -m, \ldots -2, -1, 0, 1, 2, \ldots, m \tag{2.51}$$

Table 2.5 lists the coefficients for four half band filters designed for a pass-band from 0 to $f_s/8$ and a stop band from $3f_s/8$ to $f_s/2$. Figure 2.42 compares these four half-band filter realizations. The filter with the least suppression has three non-zero coefficients increasing to nine for 72 dB suppression.

Table 2.5 The non-zero coefficients for four half band filters in Fig. 2.42 (courtesy: E.E. Janssen)

Coefficients	Suppression (dB)	Ripple (dB)
$a_m = 0.5$ $a_{m-1} = a_{m+1} = 0.2900$	20	0.8
$a_m = 0.5$ $a_{m-1} = a_{m+1} = 0.2948$ $a_{m-3} = a_{m+3} = -0.0506$	38	0.1
$a_m = 0.5$ $a_{m-1} = a_{m+1} = 0.3016$ $a_{m-3} = a_{m+3} = -0.0639$ $a_{m-5} = a_{m+5} = 0.0130$	55	0.014
$a_m = 0.5$ $a_{m-1} = a_{m+1} = 0.3054$ $a_{m-3} = a_{m+3} = -0.0723$ $a_{m-5} = a_{m+5} = 0.0206$ $a_{m-7} = a_{m+7} = -0.0037$	72	0.002

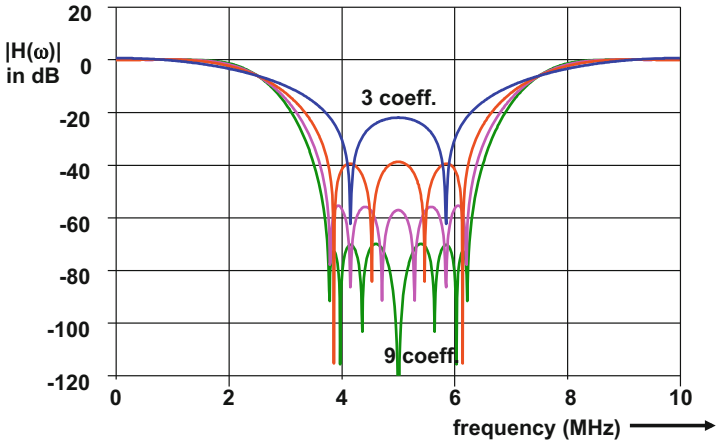


Fig. 2.42 Four half-band filters, with 3, 5, 7, and 9 non-zero coefficients (courtesy: E.E. Janssen)

In order to obtain a small-area implementation the coefficients are rounded integers. With integer filter coefficients no full multiplier circuit is needed but dedicated shift and add circuits create the weighting of the signal samples.

2.7.3 IIR Filters

A drastic solution to the hardware problem of FIR filters is the "infinite impulse response" IIR filter.

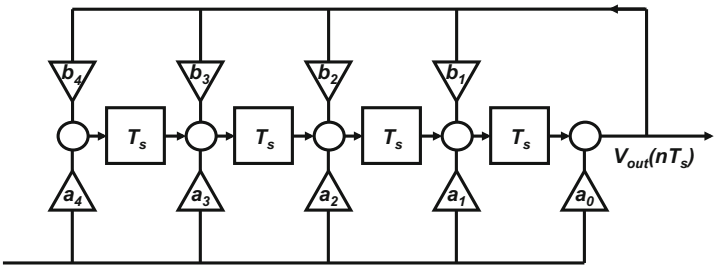


Fig. 2.43 The structure of an infinite impulse response (IIR) filter containing a feedback path from output to the summation nodes

Figure 2.43 shows the general or canonical form of a digital IIR filter. Coefficients a_0 to a_4 perform the same function as in an FIR filter. In addition the coefficients b_1 to b_4 feed the output back into the delay chain. This feedback modifies the FIR transfer. If all coefficients b_k equal zero, again an FIR filter will result.

A similarity to the RLC filter is that in both filter types the signal is circulating in the filter. In the RLC filter the signal swings between the electrical energy in the capacitor and the magnetic energy in the coil. In an IIR filter the signal circulates via the feedback path. The signal frequency in relation to the delay of the loop and the coefficients will determine whether the signal is amplified or attenuated and for how long. The transfer function of an IIR filter is (for the mapping from z-domain to frequency domain the approximation $z = e^{j\omega T_s}$ is applied):

$$H(z) = \frac{\sum_{k=0}^{k=N-1} a_k z^{-k}}{1 - \sum_{k=1}^{k=N-1} b_k z^{-k}} \Leftrightarrow H(\omega) = \frac{\sum_{k=0}^{k=N-1} a_k e^{-jk\omega T_s}}{1 - \sum_{k=1}^{k=N-1} b_k e^{-jk\omega T_s}} \tag{2.52}$$

The numerator specifies the FIR filter part, while the denominator describes the feedback path. Both are formulated as a polynomial in z^{-1}. For absolute stability (a bounded input results in a bounded output signal) the zeros of the denominator polynomial must be smaller than 1, they reside inside the unity circle of Fig. 2.34. In theory the signal will never fully extinguish in an IIR filter. In practice, a signal that experiences a feedback factor $1 - \Delta$ will pass in the order of $1/\Delta$ times through the filter. A filter with $\Delta \ll 1$ is called a resonator and resembles a high-Q RLC filter. An IIR construction where the denominator has a zero term equal to "1" will oscillate.

A sharp low-pass filter with just four delay elements as in Fig. 2.44 realizes between 2 and 4 MHz a suppression of 40 dB, which is comparable with a seventh order analog filter.

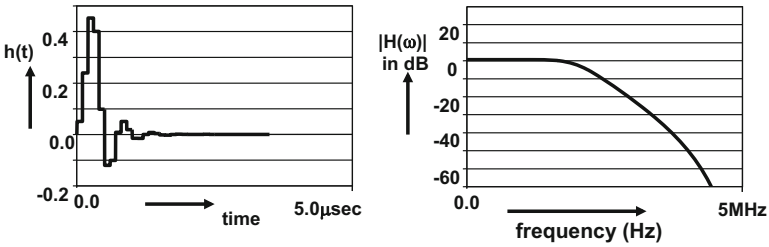


Fig. 2.44 The impulse response of this 4-tap IIR continues beyond four sample periods. The frequency response of this 4-tap filter is much steeper than the response of the 7-tap FIR filter

Table 2.6 Comparison of discrete filter realization techniques

Implementation	Switched capacitor	Semi-digital filter Sect. 7.3.7	Digital hardware
Delay	Analog	Digital	Digital
Coefficients	Analog	Analog	Digital
Most used	As IIR/resonator	As FIR	Both FIR and IIR
Noise	Accumulates in signal range	Only from coefficients	Related to word width
Tolerance	Capacitor matching	Current source matching	Unlimited
Alias-filter	Required	Depends on system requirements	Required
Power	Moderate	Output related	High
Performance	Limited by noise	Limited by noise	Limited by word width

Time-discrete filters can be realized in various implementation technologies. Table 2.6 compares three realization forms of time-discrete filters. The switched capacitor filters are mostly used in medium specification circuits. The realization is practically limited to 40–50 dB signal-to-noise levels.

Exercises

2.1. A sinusoidal signal of 33 MHz is distorted and generates second and third harmonics. It is sampled by a 32 Ms/s system. Draw the resulting spectrum.

2.2. A signal bandwidth from DC to 5 MHz must be sampled in the presence of an interferer frequency at 20 MHz. Choose an appropriate sampling rate.

2.3. An image sensor delivers a sampled-and-held signal at a fixed rate of 12 Ms/s. The succeeding digital signal processor can run at 10 MHz. Give an optimal theoretical solution. What is a (non-optimal) practical solution?

2.4. What is a stroboscope? Explain its relation to sampling.

2.5. Must the choice for a chopping frequency obey the Nyquist criterion?

2.6. Set up a circuit where the signal is stored as a current in a coil. What is the meaning of $i_{noise} = \sqrt{kT/L}$?

2.7. In Example 11.6 the available equipment can only measure signals up to 5 MHz. What can be done in order to measure the harmonic distortion of the analog-to-digital converter at roughly 5 GHz?

2.8. The signal energy of the luminance (black-and-white) signal of a television system is concentrated around multiples of the line frequency (15,625 Hz). Although the total television luminance signal is 3 MHz wide, a sample rate of around 4 Ms/s is possible. Give an exact sample rate. Why will this sampling not work for a complete television signal with color components added?

2.9. An audio system produces samples at a rate of 44.1 ks/s. With a maximum audio signal of −6 dB of the full-scale between 10 and 20 kHz, propose an alias filter that will reduce the frequency components over 20 kHz to a level below −90 dB.

2.10. How much SNR can be obtained if a signal of 10 MHz is sampled with a sample rate of 80 Ms/s with 5 ps_{rms} jitter. What happens with the SNR if the sample speed is increased to 310 Ms/s at the same jitter specification. Compare also the SNR in a bandwidth between 9 and 11 MHz.

2.11. Design a half band filter with 19 non-zero coefficients to get a pass-band stop band ratio of 100 dB. Use a computer program.

2.12. An analog-to-digital converter is sampling at a frequency of just 2.5× the bandwidth. Due to the large spread in passive components, the problem of alias filtering is addressed by placing before the converter a time-discrete filter running at twice the sample rate and before that time-discrete filter a time-continuous filter. Is this a viable approach? There are twice as many samples coming out-of the filter then the converter can handle. Is there a problem?

2.13. Make a proposal for the implementation of the filters in the previous exercise if the bandwidth is 400 kHz and a attenuation of better than 60 dB must be reached starting from 500 kHz.

2.14. In Example 2.14 the second capacitor is twice the size of the first: $C_2 = 2C_1$. Will the signal-to-noise ratio remain the same?

2.15. A sinusoidal signal of 33 MHz is distorted and generates second and third harmonics. It is sampled by a 32 Ms/s or a 132 Ms/s system. Draw the resulting spectra. What sample rate do you favor?

2.16. A signal source delivers a signal that consists of three components: at 3, 4, and 5 MHz. The signal is processed by a sampling system with an unknown sample rate. The output contains in the 0–0.5 MHz band only frequencies at 0.27, 0.36, 0.45, and 0.46 MHz. What sampling frequency was used? Complete the spectrum till 1 MHz.

2.17. An RF oscillator at 2.45 GHz contains harmonic distortion products at 2x and 3x time the oscillation frequency. The available spectrum analyzer can measure up to 10 MHz, but has a 10 GHz input bandwidth sampling circuit with variable sampling rate up to 10 GS/s. Advice how to measure the harmonic distortion.

2.18. A 2 MHz signal is sampled by a 100 Ms/s clock with 10 ps_{rms} jitter. In the digital domain the band of interest is limited to DC-5 MHz, Calculate the SNR. The digital circuits repeat a process every 10 ms and this is the cause of the 10 ps jitter. What is the resulting spectrum in DC -5 MHz?

Chapter 3
Sample and Hold

3.1 Track-and-Hold and Sample-and-Hold Circuits

In Chap. 2 the theory of sampling is described. A designer of a complex system will try to concentrate the limitations of the sampling process and the optimization of the function into one circuit block of the system. Often this function is realized as a "track-and-hold" (T&H) circuit that creates a stable signal for a part of the sample period of the ADC. The most elementary T&H circuit consists of a switch and a capacitor, Fig. 3.1. During the conducting phase of the switch, the signal on the capacitor follows the input signal, while in the isolating phase (the hold phase) the signal value remains fixed at its value at the moment of opening the switch. This moment is the theoretical sampling point in time.

Two T&H circuits connected in cascade form a sample-and-hold circuit (S&H). The second T&H circuit is triggered by an inverted or delayed sampling signal. Figure 3.2 shows the input signal and the output of a T&H and an S&H circuit during track-and-hold operation. An S&H circuit will hold the signal over the full period of the sampling clock. This is a necessary condition for time-continuous restoration when the sample-and-hold function is at the end of the analog-to-digital-to-analog signal chain. And in certain converter architectures a sample available for a full sample period allows more processing to be done.

Track-and-hold and sample-and-hold circuits are used for performing the sampling operation on the analog input signal at a sample moment in an analog-to-digital converter. The T&H circuit keeps the sampled signal at its level for a time period thus allowing repeated use of the signal during the analog-to-digital conversion, see Fig. 3.3 (left). The output value of the track or sample-and-hold remains, however, associated with the original sampling moment. The fact that the output value is internally used at slightly delayed time moments will not create any effect in signal transfer characteristics, as long as in the processing the sample value is assigned to the original sampling moment.

M. Pelgrom, *Analog-to-Digital Conversion*, DOI 10.1007/978-3-319-44971-5_3

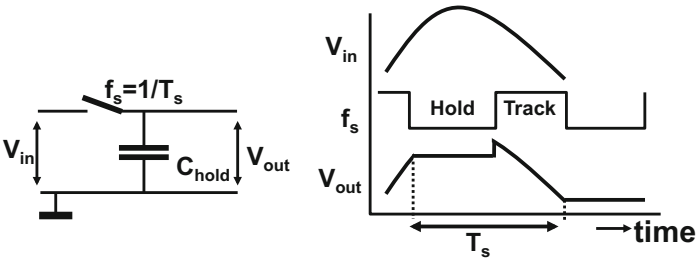


Fig. 3.1 A switch and a storage element form a track-and-hold circuit

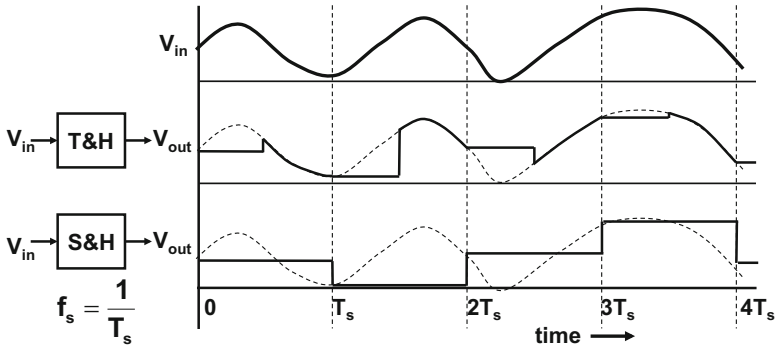


Fig. 3.2 The input signal (*above*) is tracked by a track-and-hold circuit during one phase of the clock period and held during the other phase. A sample-and-hold circuit holds the signal during the full period of the clock signal

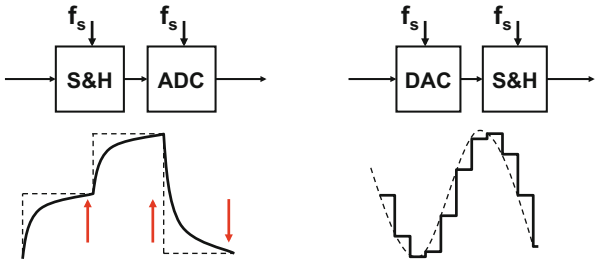


Fig. 3.3 A sample-and-hold circuit is used as an input sampler for an analog-to-digital converter or as a de-glitch circuit in an digital-to-analog converter

As the output value of the T&H or S&H circuit is used only at specific time moments, some implementation compromises may be acceptable for obtaining a high quality output signal. For example, slewing during the initial phase of the settling will not affect the overall performance as long as a stable value is reached at the required time moment(s).

3.2 Artifacts

The sample-and-hold circuit has to fulfill the requirements of maximum sampling speed and bandwidth in the analog domain as well as in the sampled data domain. In order to achieve maximum signal-to-noise ratio the T&H circuit has to handle the maximum amplitude possible. The speed and amplitude specifications make that many designers regard the T&H circuit as one of the most critical components in realizing an optimum conversion performance.

The specification of the performance of a T&H circuit consists of the standard analog ingredients: distortion (THD), signal-to-noise ratio (SNR), power consumption (P), etc. Next to these standard requirements, a number of specific specifications exists for a T&H circuit, see Fig. 3.4:

- Cross-talk from the switching pulse may cause an instantaneous drop[1] in the voltage of the hold capacitor, called the "pedestal step." Another contribution to the pedestal step comes from removal of the charge Q_{gate} that forms the conductive layer of the MOS switch or the base charge in a bipolar transistor. This charge will flow back into the signal source and into the hold capacitor. In the ideal circuit of Fig. 3.4 with an NMOS switch and perfectly symmetrical charge split:

$$V_{pedestal} = \frac{C_{ped}\Delta V_{gate} + Q_{gate}/2}{C_{hold}} = -\frac{WC_{olap}V_{DD} + WLC_{ox}(V_{DD} - V_{in} - V_{T,N})/2}{C_{hold}} \tag{3.1}$$

  Typically most of C_{ped} is gate overlap capacitance and proportional to the transistor width W and the specific overlap capacitance per micron: C_{olap}. A

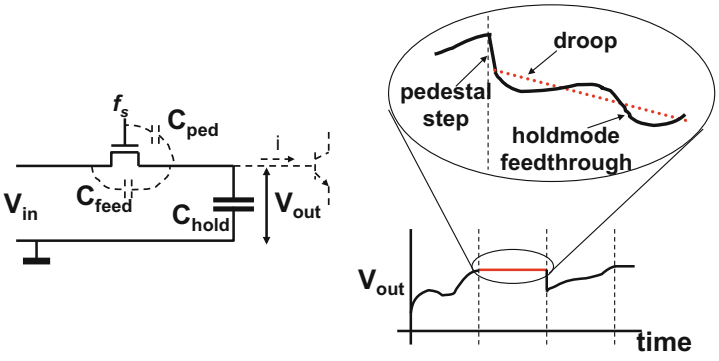


Fig. 3.4 Three artifacts in a sample-and-hold: droop, hold-mode feed-through, and pedestal step

[1]For convenience reasons the switch is assumed to be implemented as an NMOS transistor, unless otherwise stated. Conduction takes place with a positive gate voltage.

closer look at the charge injection shows that there is a slight amplification at the moment of sampling. Looking just at the signal-dependent part of the output voltage:

$$V_{hold} = V_{in} + \frac{Q_{gate}(V_{in})}{2C_{hold}} = V_{in}(1 + \frac{WLC_{ox}}{2C_{hold}}) + V_{DC} \tag{3.2}$$

This effect can jeopardize the performance of pipeline and algorithmic based converters where an exactly multiplication factor is required, see Sects. 8.4 and 8.7. For example, a switch with dimensions 10/0.06 µm in a 65-nm process will show a gate capacitance of 8 fF, Table 4. With a hold capacitor of 400 fF, this would result in a 1.01× amplification.

In advanced technologies with small gate-lengths the channel contains much less charge and therefore this amplification is less an issue.

- The charge will not always split evenly. The typical rise/fall time of a switching pulse is 50 fs in 65-nm CMOS. At these speeds the channel cannot be considered a single element, its internal time constant is of the same order of magnitude. The charge dump must be analyzed as charge moving along a transmission line to the terminals of the switch, see Fig. 3.5. Depending on the impedances on both sides of the switch and the switching speed of the sample pulse, significant deviations can occur in the splitting accuracy. This leads to unexpected signal components in the voltage over the hold capacitor. In some cases the impedance can be made more equal by adding a capacitor on the input side of the T&H circuit.

  Compensation of the pedestal by means of half-sized transistors is possible. In a practical design of Fig. 3.6 the dummy switches are single devices while the pass transistor consists of two parallel transistors for optimum cancellation. The control of the sample pulse edges requires careful balancing to a level lower than the rise time. If the "clk" pulse switches off before the inverse pulse becomes active, there will be a lot of switching glitches. If the inverse pulses become active while the "clk" pulse is still active, the compensation charge will be supplied by the source and the whole method is ineffective. The occurrence of these two situations partly depends on the level of the input signal.

  The dummy switches remove the constant part of the pedestal step. Mismatch between the transistors creates a random charge component. This random part is

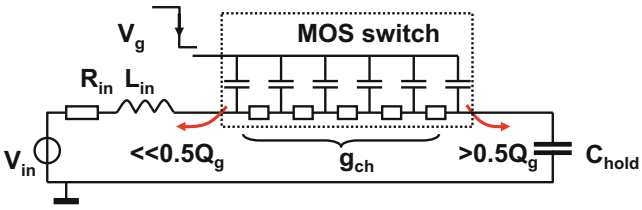


Fig. 3.5 The channel charge in an MOS device acts as a transmission line when the transistor is switched off. If the impedances on either side differ, the charge splitting is asymmetrical

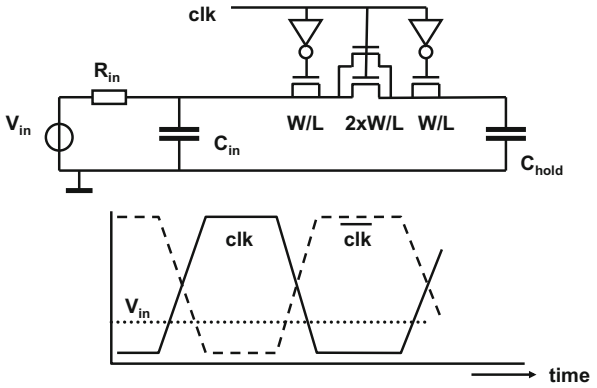
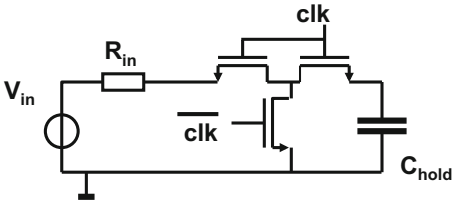


Fig. 3.6 The inherent charge splitting can be compensated by dummy switches

Fig. 3.7 The T-switch reduces the signal feed-through



increased as the random variation in the channel charge of the switch transistor adds to the random variation of the transistors providing the compensating charge.

- During the hold phase of the track-and-hold, charge can leak from the hold capacitor and the signal will show "droop." In a bipolar design this leakage is caused by the base current of the next circuit. This effect results in the specification of a minimum sample rate of, e.g., a few hundred ksamples/s. In deep submicron processes gates may become so leaky that again droop becomes a relevant parameter.

$$V_{droop} = -\frac{I_{leak} T_{hold}}{C_{hold}} \tag{3.3}$$

- The hold-mode feed-through describes the parasitic transfer from signal source to the output of the circuit in hold mode. Next to trivial reasons, such as unintended coupling via wiring or power supply, some residual coupling may occur due to the source–drain capacitor in an MOS switch. Although normally hold-mode feed-through is minimum in integrated solutions, a T-switch may further reduce this effect, Fig. 3.7. This switch is a series connection of two MOS devices both clocked with the sample pulse. A third device connects the common source–drain terminal of the series switches to ground. This device is clocked in anti-phase and creates a strong attenuation of the feed-through signal.

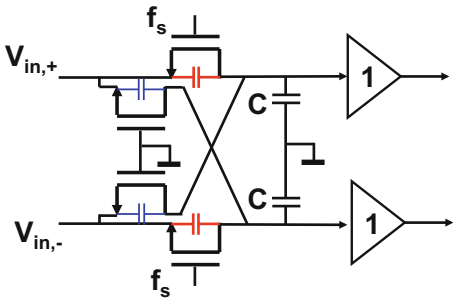


Fig. 3.8 The differential sampling scheme used in Fig. 9.37 compensates the feed-through component of the short switch transistors by means of identical feed-through capacitors of the always-off transistors

A pseudo-differential solution for the T&H was implemented in Fig. 3.8 and applied in the time-interleaved converter discussed with Fig. 9.37. The feed-through signals coming via the source–drain capacitances of the very short switches are compensated by always off cross-coupled devices.

- The aperture time is the time period in which the sample value is determined. Normally this is a relatively small time period in which the sample pulse makes the transition to switch off. During this clock edge still some tracking of the signal can continue. This will add a small additional delay to the actual sampling moment with respect to the start of the switch-off process. More devastating are random changes in aperture time. Aperture jitter is equivalent to the jitter discussed in Sect. 2.6.1.

  Aperture distortion is introduced when the slew rate of the signal is comparable to the transition time of the switch. A signal change that follows the edge of the switching pulse postpones the sampling moment, while a signal change in opposite direction of the switching pulse will advance the effective sampling moment.

  One analog-to-digital architecture has a deliberately long aperture time: the dual-slope converter, see Sect. 8.8. Here the signal is integrated during the aperture time. The effect on the transfer characteristic is a filtering of high-frequency components (noise), which is a desired effect in applications of this converter.

Example 3.1. What is the minimum frequency on which a track-and-hold circuit can operate, if the hold capacitor of 1 pF experiences a leakage current of 1 nA? The output must be accurate within 1 mV.

Solution. The voltage on the hold capacitor will see a voltage change of $\Delta V_{Chold} = I_{leakage} \times T_{hold}/C_{hold}$ with the above numbers $T_{hold} \leq 1\,\mu s$. If the hold time is dominant, the minimum sample rate is 1 Ms/s.

3.3 Capacitor and Switch Implementations

3.3.1 Capacitor

The sample switch and the hold capacitor are the prime components of a track-and-hold or sample-and-hold circuit. The value of the hold capacitor for signal-to-noise ratios in excess of 40–50 dB is determined by kT/C_{hold} noise.

The distortion requirements on the capacitor depend on the way it is used. If there is ample settling time and the capacitor is part of a voltage buffer, the linearity requirements are low. A simple capacitor with the plate with most parasitic components (the bottom plate) connected to ground can be read out as a voltage buffer and can tolerate some voltage dependence of the capacitance.

A capacitor that is used as a voltage-to-charge converter, as in most switched capacitor implementations, must have a perfect linear voltage–charge transfer. Now the capacitor value is essential for the operation of the circuit. The requirements on its linearity are set by the overall circuit specification. See Sect. 5.2.2 for a further discussion of capacitor properties and a description of capacitors.

3.3.2 Switch Topologies

The track-and-hold switch is characterized by its on/off impedances. In a T&H circuit the on-resistance must be small and constant and the off-impedance must be infinite. Obviously the conductive switch and capacitor in the T&H circuit form a time constant $\tau = R_{on}C_{hold}$. This time constant must be low enough to approximate the input signal with sufficient accuracy. Moreover, the previous sample must be removed avoiding that samples influence each other. Both processes are determined by an exponential function. The error voltage for both processes is

$$V_{error} = (V_{sample,i} - V_{sample,i-1})e^{-T_{switch}/\tau} \tag{3.4}$$

where T_{switch} is the period where the switch is conducting. For example, if the error voltage should be less than 0.1 % of the previous sample value, then $\tau < T_{switch}/7$ ($e^{-7} = 10^{-3}$). With a duty cycle of the sample rate of 50 % ($T_{switch} = T_s/2$) the required bandwidth of the switch resistance and the hold capacitor is $BW > 2.5f_s$. In case of more complex configurations as in the next paragraphs, the settling time of the overall structure may be dominated by the opamp unity-gain frequency.

When a single MOS transistor is used as a switch, the on-state conductivity of the channel depends on the gate-to-source voltage minus the threshold voltage:

$$R_{on,NMOS} = \frac{1}{(W/L)_N \beta_{N\square}(V_{DD} - V_{in} - V_{T,N})}$$

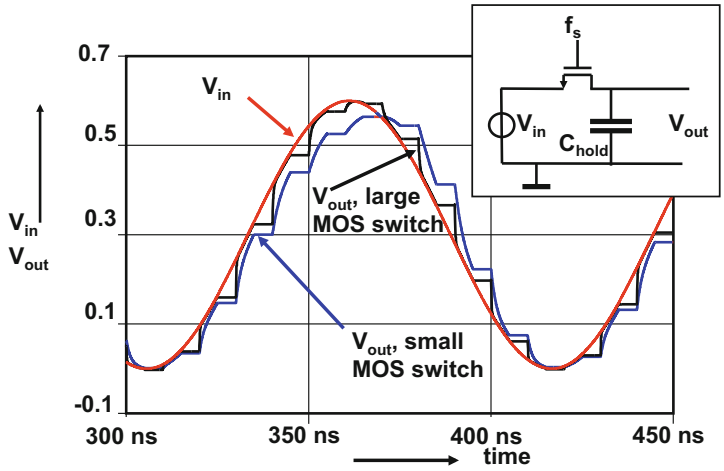


Fig. 3.9 The resistance of a small NMOS switch causes significant distortion in the hold signal

$$R_{on,PMOS} = \frac{1}{(W/L)_P \beta_{P\square}(V_{in} - |V_{T,P}|)} \tag{3.5}$$

The NMOS switch is conductive if the input voltage is lower than the gate voltage minus the threshold voltage. The maximum gate voltage often equals the power supply voltage. The PMOS switch is conductive with input voltages that exceed a threshold voltage above the gate voltage (mostly zero or ground level). At low supply voltages and large signal excursions the voltage dependent resistance of the switch can lead to aperture time differences causing distortion. Figure 3.9 shows a track-and-hold circuit simulation in a 1.2 V 90-nm CMOS process. The sinusoidal input signal is sampled with 100 Ms/s on a 10 pF capacitor. The simulation with a wide switching transistor (50/0.1 µm) shows nearly no artifacts. However, the tenfold narrower transistor shows clearly that the delay between input and output increases at higher signal voltages. At those higher signal voltages there is less channel charge in the NMOS transistor and the $R_{on}C_{hold}$-time increases. The net effect of this signal-dependent aperture delay time is distortion.

With a simple approximation the magnitude of the distortion is estimated. Assume that the total variation of the resistor with nominal value R_0 over the signal range is ΔR and this resistance change is approximately linearly dependent on the signal:

$$R(V_{in}(t)) = R_0 + \frac{V_{in}(t)}{V_{in,peak-peak}} \Delta R$$

With an input signal $V_{in}(t) = 0.5\,\mathrm{V}_{in,peak-peak} \sin(\omega t)$ the current is mainly determined by the capacitor: $I(t) \approx \omega C 0.5\,\mathrm{V}_{in,peak-peak} \cos(\omega t)$. The voltage drop

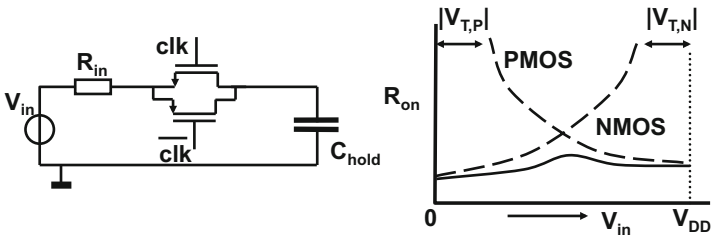


Fig. 3.10 The complementary PMOS-NMOS switch. Also known as "transmission gate"

over the resistor is composed of linear and second order terms. This distortion amplitude over the resistor is equal to the second order term that will appear over the capacitor. The HD2 is then found as:

$$\text{HD2} = \frac{\omega \Delta RC}{4} \tag{3.6}$$

Some fine tuning in a simulator is needed as the switch resistance also shows higher-order artifacts.

This signal-dependent impedance is the main problem with a simple switch controlled by fixed voltages. Also in more complicated switched-capacitor structures there is always one critical switch in this respect.

Figure 3.10 shows a popular implementation of a solution towards obtaining a constant-resistance switch. An NMOS transistor and a parallel connected PMOS transistor compensate each other's poor conductivity regions. The overall resistivity of the switch over the input voltage range is considerably more flat. The use of two transistors implies that the controlling clock pulses come exactly at the right position in time. If the PMOS transistor switches slightly before the NMOS transistor (or vice-versa) again a form of sampling time difference will occur between the upper half of the signal and the lower part. The switch resistance varies less than of a single transistor switch, but some resistance modulation with the signal amplitude is still present.

At extremely low supply voltages where $V_{T,N} + |V_{T,P}| < V_{DD}$, an input region exists where none of the two transistors will conduct. In [34, 35] the switched operational amplifier technique is presented as a solution for this case. The functionality of the switch is implemented in the driver transistors of the opamp output stage. In case the opamp output stage switches to a high impedance, an additional switch is needed to connect the hold capacitor to a stable voltage. The switching operation in the opamp itself causes internal disturbances. An equivalent or better performance to the standard techniques is only reached in cases with very low power supplies. In many designs a better performance is reached with a bootstrapping technique.

Example 3.2. What will happen to a 10 MHz sine wave if the PMOS switch in Fig. 3.10 switches off 50 ps later than the NMOS switch?

Solution. A delay between the two switches will result in a time shift of the upper and lower part of the signal. In theory a complex signal processor can correct, but in practice the resulting sample sequence will show a distorted behavior. As the error is in the time domain, and resembles jitter behavior, a first order calculation will give a rough estimate of the resulting distortion. Figure 2.26 indicates that a time error Δt results in an amplitude error of $\Delta A = \omega \hat{A} \cos(\omega n T_s) \Delta t(n T_s)$. At mid range the amplitude error relative to the peak amplitude is $(2\pi f \Delta t) \approx -50$ dB.

3.3.3 High Voltages

The obvious solution to too low gate voltages is to use internally generated high voltages. Various techniques exist to multiply voltages. These techniques are either based on inductive (buck and boost converters) or capacitive multiplication schemes. In the context of sampling the required energies are limited and only capacitive schemes are selected using the available components in a CMOS technology. Figure 3.11 shows two basic voltage multiplication techniques. On the left side the capacitors are connected in series and driven by two complementary pulses. The diodes (or switches) charge the intermediate nodes to the swing of the driving pulses. If all parasitic elements are under control this scheme can generate output voltages in excess of the individual capacitor breakdown voltage. A limitation of this scheme is that the effectiveness depends on the ratio between the series connection of capacitors and the parasitic capacitors.

Dickson [36] proposed a parallel technique that allows a more effective voltage multiplication in the presence of parasitic capacitive load on the nodes. The parasitics associated with the bottom-plate of the capacitor are charged by the pulse generator. Again the circuit is driven by complementary pulses and every node is charged to a voltage of roughly:

$$V_n = n\left(\frac{C}{C + C_{par}} V_{clock} - V_{diode}\right) \tag{3.7}$$

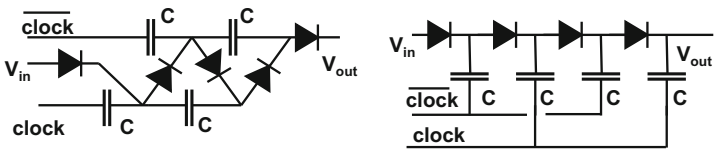


Fig. 3.11 Two techniques to generate a high voltage. *Left*: the series connected Cockcroft–Walton voltage multiplier (1932). *Right*: the parallel connected Dickson multiplier [36]

where n reflects the number of stages.

This formula points to the main efficiency limiting effects in many voltage multiplication circuits: the ratio of the parasitic capacitance to the driving capacitance and the voltage loss over the rectifier element. Nevertheless these circuits find application in, e.g., EEPROMs.

3.3.4 CMOS Bootstrap Techniques

A popular technique to circumvent most of the above problems and to keep the on-resistance of a sampling switch constant is the class of bootstrapping[2] circuits, Fig. 3.12. In these schemes the effective drive voltage is increased beyond the power supply limits. In the era of enhancement/depletion NMOS technology Knepper [37] used a bootstrap circuit to solve the problem of driving an NMOS transistor with gate voltages higher than the power supply. This principle has been implemented in CMOS in various forms, mostly with only one or two capacitors. These circuits aim to solve two problems associated with the one-transistor switch: the limited input range due to the threshold voltage and the switch resistance variation.

Figures 3.12 and 3.13 show the general idea of bootstrapping the drive voltage of a sampling switch. A capacitor with a value of typically ten times the gate capacitance of the switch is charged to the power supply during the hold phase of the T&H circuit. In the track phase this capacitor is connected between the input voltage and the gate of the transistor. Now the input voltage of the T&H serves to push the capacitor top plate voltage to levels beyond the supply. The gate of the

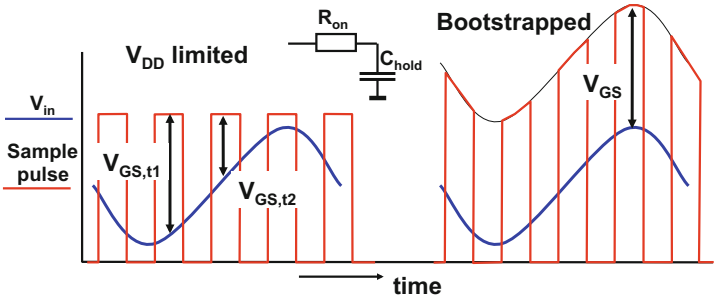


Fig. 3.12 The drive voltage of the switch varies if the power supply voltage is fixed (*left*) and consequently the resistance that charges the hold capacitor varies as well. If the switch is bootstrapped, ideally the drive voltage remains constant over the input range

[2] A relation seems likely with the tales on Baron von Mŭnchhausen, who pulled himself up by his bootstraps, and "booting" of computers.

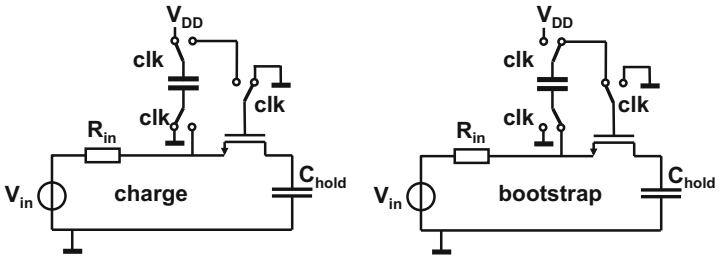


Fig. 3.13 The bootstrapping principle

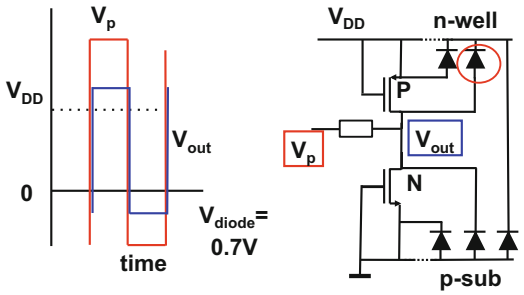


Fig. 3.14 A high-swing pulse V_P is applied to an NMOS and PMOS transistor. The diode formed by the p-type drain and the n-type well doping will conduct and level off the maximum voltage swing. The same will happen on the negative side if a negative-going pulse opens the diode on the NMOS drain

sampling transistor is now biased at a constant voltage with respect to its source and drain, independent of the input level.

Several issues must be considered when implementing a bootstrap circuit. Generally a foundry will not allow more voltage drop over any combination of drain, source, and gate terminals of one transistor than V_{DD}. Using a high positive voltage swing on PMOS transistor will open the p-type drain to n-well diode and simply limit the maximum level, see Fig. 3.14. The associated current will flow into the substrate, losing charge and eventually causing latch-up problems.

Abbo and Gray [38] proposed a bootstrap circuit that circumvents most issues. In Fig. 3.15 transistors T_1 and T_2 implement a clock voltage multiplier and operate as the diodes in Fig. 3.11.

The analysis of the circuit starts by assuming all capacitors are discharged ($V_{C1} = V_{C2} = V_{CB} = 0$. If the clock signal f_s goes high, the source of T_1 is pulled low, while the gate is pushed up via capacitor C_2. Note that ideally the voltage over this capacitor will not change as the isolating transistor T_2 prevent its top plate to discharge. During this clock edge T_1 will (partially) charge the capacitor C_1. In the next clock cycle the same will happen with the capacitor C_2. In the following clock cycles the extra charge on one capacitor will increase the drive voltage for charging the other capacitor until both are charged to the full power supply voltage V_{DD}. After fully charging of both capacitors one of the transistors T_1 and T_2 will experience a gate voltage of almost $2V_{DD}$ and source and drain voltages of V_{DD}, while the other

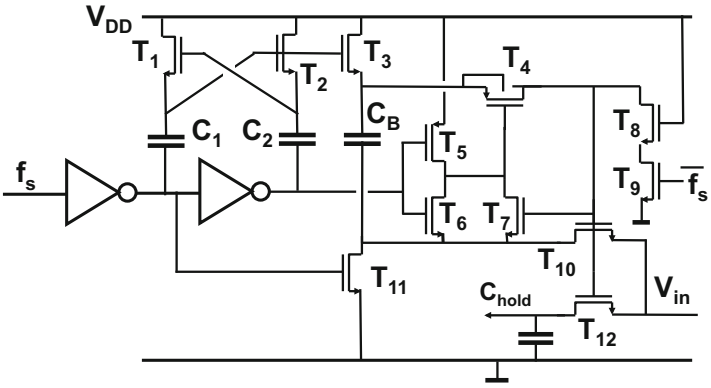


Fig. 3.15 The bootstrap circuit proposed by Abbo and Gray [38]

transistor will isolate because its drain and gate voltages are at V_{DD} and its source[3] voltage is at almost $2V_{DD}$. The maximum voltage will never reach exactly $2V_{DD}$ because of the parasitic capacitive load on the top plate of the capacitors C_1 and C_2.

Parallel to T_2, transistors T_3 and T_{11} will charge the bootstrap capacitor C_B.

At the onset of the tracking phase (the clock f_s goes high) T_3 and T_{11} switch off and are isolating. C_B was charged to V_{DD} in the previous cycles and acts as a local battery for T_4 to T_7. T_6 connects the gate of T_4 to the low side of C_B. Charge in C_B will pass via T_4 to the central bootstrap node that connects to T_{10} and the T&H switch transistor T_{12}. Via T_{10} the input voltage now starts pushing the bottom plate of C_B and the bootstrap action starts.

Sampling occurs when the clock goes low, T_8 and T_9 are activated. These transistors rapidly discharge the central bootstrap node and the circuit goes in the hold phase. The delay T_{delay} between the moment T_9 starts conducting and the actual off-switching of T_{12} is determined by the capacitance $C_{G,12}$ connected to the gate of T_{12}, the fixed turn-on voltage V_{on} of T_{12}, and the saturation current $I_{Dsat,T9}$ of T_9. As these three elements are in first order identical for every sample value, the delay $T_{delay} = V_{on}C_{G,12}/I_{Dsat,T9}$ will be constant. The skew and variation on the sample pulse remain the dominant source for jitter.

The requirements for process reliability demand that voltages between any two nodes (drain, gate, and source) of a transistor remain at, or below, V_{DD}. After some clock cycles the gate, source, and drain voltages of T_1, T_2, and T_3 will swing between V_{DD} and almost $2V_{DD}$. The drain–source voltage of T_9 cannot exceed the voltage limits, as T_8 serves as a protection device: in this way the maximum voltage on C_B is divided over both transistors T_8 and T_9. This bootstrap circuit violates the reliability demands during transitions, e.g., if the voltage on the hold capacitor is

[3]The word "source" is used for the device terminal with the arrow in the schematic, irrespective of the voltage.

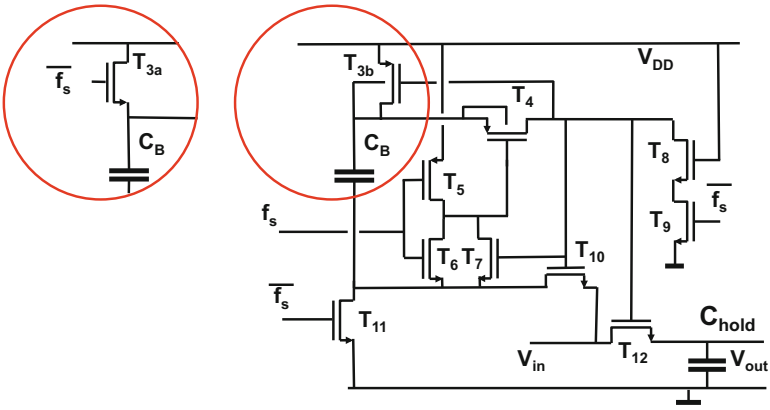


Fig. 3.16 A simplified bootstrapping circuit modified from [39, 164]

near ground and charging is slow, the gate-to-drain voltage of T_{12} can exceed V_{DD}. Still reliability people insist to prove that all devices are protected from over-voltage under all circumstances.

The previous bootstrap circuit can be simplified, see Fig. 3.16. The charge pump circuit has been removed. Transistor T_{3a} is connected directly to the inverse f_s Fig. 3.16 (insert) or is replaced into a PMOS device T_{3b}. With an NMOS transistor the maximum voltage on the bootstrap capacitor is $V_{DD} - V_{T,T3a}$ and consequently the bootstrap circuit generates less overdrive than in Fig. 3.15.

When the PMOS transistor is used, the n-well of this device must be connected to the highest voltage in the circuit, in this case the bootstrapped voltage, to prevent its drain diode to conduct. Extra loading of the bootstrap capacitor C_B and less drive voltage for the sample switch will be the result. Some authors leave out T_7 and argue that the gate voltage of T_4 will remain. This is an unnecessary risk.

The above circuits bootstrap the drive voltage to almost $2 \times V_{DD}$. When sufficient power supply voltage is available the need to go to $2V_{DD}$ or go above V_{DD} is less stringent, yet it is advantageous to use a bootstrap technique in order to keep the on-resistance constant.

3.3.5 Bottom Plate Sampling

In the previous circuits a signal-dependent channel charge influences the sampling process. "Bottom-plate" sampling introduces a second transistor switch on the grounded side of the capacitor, see Fig. 3.17 (top). This switch uses the full power supply range for a maximum gate drive and is not impaired by input signal variations as it connects to ground. The bottom-plate switch is turned off before the original sampling switch, thereby isolating the capacitor. A capacitor isolated on one side

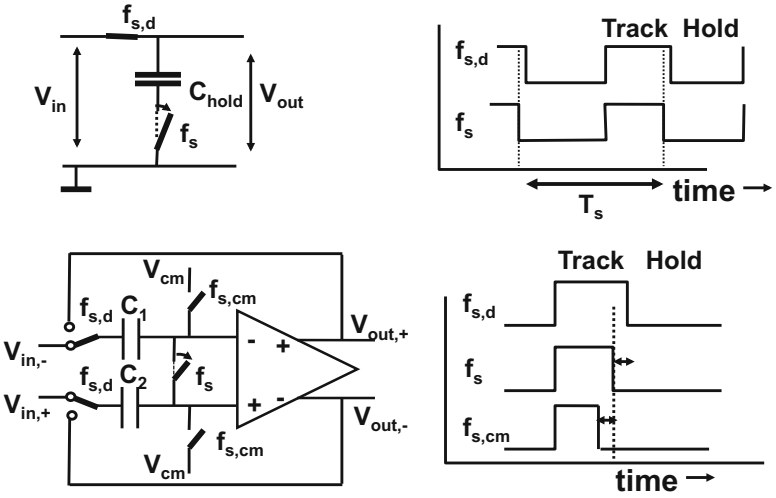


Fig. 3.17 *Top*: principle of bottom-plate sampling. The switch to the bottom plate is driven by f_s and opens before $f_{s,d}$ disconnects the top plate. *Bottom*: differential variant of bottom plate sampling. The switches in both schemes are drawn at the moment the bottom plate disconnects via f_s [40]

cannot exchange charge on the other side. As long as the top side switch turns off before the charging of the parasitic capacitors connected to the bottom plate becomes a problem, this bottom-plate sampling will partly reduce the modulation problems with the sample switch.

Figure 3.17 (bottom) shows the bottom plate sampling technique in a differential circuit. In track mode the input terminals are connected to the left-hand sides of C_1 and C_2. The right-hand sides are connected together by means of the bottom-plate sample switch. Moreover two common mode switches are used to ensure that the floating opamp input nodes are set at the right bias. These switches are opened first, followed by the bottom-plate sample switch. Now the capacitors can no longer exchange charge, assuming the inputs of the amplifier are ideal. The sequence is finalized by disconnecting the capacitors from the input voltage and closing the feedback loop of the opamp and the capacitors.

3.4 Track-and-Hold Circuit Topologies

3.4.1 Buffering the Hold Capacitor

The voltage on the hold capacitor must be buffered, Fig. 3.18, before any succeeding operation can be performed. For low-resolution and high-speed operation open-loop amplifiers such as source followers [39] and degenerated differential pairs are used.

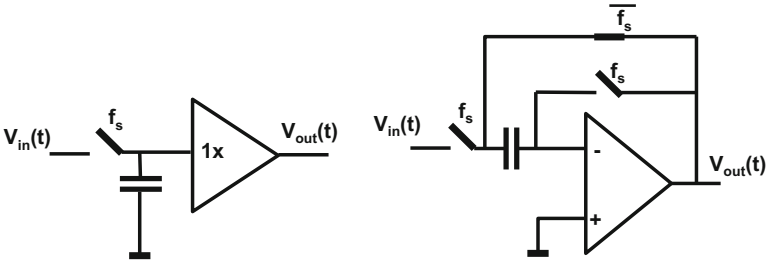


Fig. 3.18 Buffering the capacitor with a 1x amplifier or an opamp

The inherent speed advantage must be traded off versus the non-linearity. Still an 8–10 bit performance level is possible.

Higher linearity is obtained in a feedback topology with operational (transconductance) amplifiers. These opamps buffer the capacitor voltage and often perform additional functions as subtraction or amplification for the specific analog-to-digital architecture. This results in specific demands for each conversion topology.

Some general remarks:

- Any buffer will load the T&H capacitor. In a lot of topologies this loading is constant and limited. If a source follower circuit is applied the hold capacitance/impedance becomes a complex function. An interesting analysis is found in [39, Appendix].
- The signal range of the sampling switch, the input range of the buffer, and the output range of the buffer are often linked. The input range and output range of a source follower T&H differ by the threshold and drive voltage. In combination with a switch transistor of the same type, e.g., as in Fig. 3.20, the signal range is limited to $V_{DD} - 2V_T - 2V_{drive}$.

  In case an opamp-feedback topology is used, the voltage range of the switches must equal the output range and the input range, to avoid loss in voltage swing. Input and output ranges of the T&H depend on the opamp topology. An NMOS switch combined with the same transistor-type input stage will lose signal range; both at the ground and power supply side. This is not an optimum choice for low power supply application. Therefore the design of a track-and-hold circuit is best started by identifying the range requirements.
- If the buffer is constructed as a unity feedback opamp, the minimum DC amplification of the opamp is given by:

$$A_{DC} > \frac{1}{\epsilon} = 2^N \tag{3.8}$$

  where ϵ is the remaining settling error. The implicit assumption is that ϵ is unknown and is better kept a low level. In fact, in most opamp topologies the larger part of ϵ is linearly dependent on the signal and would result in a minor

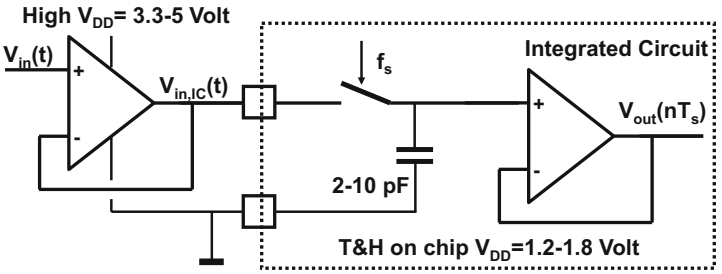


Fig. 3.19 An off-chip buffer drives the switch which is directly connected to the bondpad

overall gain error. In most analog-to-digital converters this is quite acceptable, however in, e.g., a 1-bit pipeline converter an exact gain is essential for accuracy.

- The settling speed depends on the first moment in time when an accurate output signal of the T&H is needed. In order to reach a settling level of 2π time constants within one sample period ($e^{2\pi} = 527 \approx 2^9$), the unity-gain frequency of the T&H buffer must be equal to the sample frequency. If only a fraction of the sample pulse is available for settling, the settling requirement will require far higher opamp bandwidths, e.g., UGBW$> 3f_s$.

3.4.2 Basic T&H Configurations

A simple implementation of a T&H is just a switch and capacitor connected to the input terminal of the IC. The buffers that drive this track-and-hold switch and capacitor have to fulfill the maximum specifications in terms of bandwidth, distortion, and timing accuracy. With the low supply voltages in modern IC processes these demands are difficult to realize. In some system partitioning solutions this problem is circumvented by using an off-chip driver, see Fig. 3.19. The off-chip driver uses higher supply voltages to meet the specifications, moreover the inherent S&H power consumption does not add to the power in the data sheet.

A disadvantage of this solution is that the bondpad normally is connected to a protection circuit. A series resistor and some diffusion capacitor load the input terminal and may limit the achievable bandwidth and distortion. The input signal swing can modulate the sample switch resistance, mostly a bootstrapped switch is needed. But with on terminal of the sample switch terminal connected to the outside world, a serious reliability problem may exist.

Also the on-chip handling of the signal has a number of disadvantages. Buffering the capacitor voltage requires a sufficiently large input range. The buffer needs to be designed with a sufficiently high common mode rejection ratio and any offset or 1/f noise in the buffer adds up.

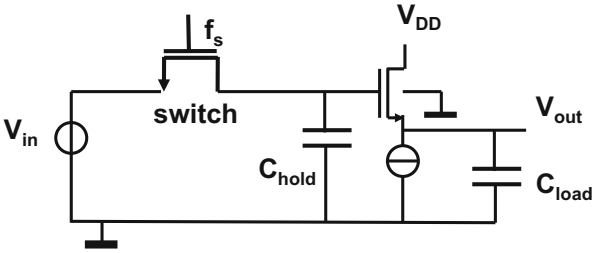


Fig. 3.20 A T&H buffered with a source follower

Example 3.3. Design a track-and-hold circuit based on Fig. 3.20 for 60 dB performance on an input signal of 0.4 V_{pp} at a bias voltage of your choice in a frequency range 0–1 GHz.

The CMOS process has a nominal power supply of 1.2 V and a minimum gate-length of 0.1 μm. The NMOST has a threshold voltage given by $V_T = 0.1 + 0.1\sqrt{|V_{BS}| + 0.6}$ volt where V_{BS} is the voltage between the source and the bulk(=ground), $\beta_\square = 300\,\mu A/V^2$. The current source supplies 1 mA and operates ideally down to $V_{out} = 0.1$ V, below 0.1 V it acts as a resistor of 100 Ω. $C_{load} = 1$ pF.

Solution. The signal space is limited: for high-voltage signals the power supply voltage minus the threshold voltage $V_T = 0.23$ V limits highest input voltage to $V_{in,max} = 0.97$ V. Lowest input is set by the current source: 0.1 V, $V_T = 0.18$ V, so $V_{in,min} = 0.28$ V. So an optimal bias level is found in the middle: 0.62 V where both transistor are biased at a drive of $V_{GS} - V_T = 0.15$ V.

The capacitor value is based on the required noise performance. With an overall 70 dB budget, the contribution of the kT/C noise is chosen at 64 dB:

$$\text{SNR} = \frac{V_{in,rms}^2}{kT/C_{hold}} = \frac{(0.4/2\sqrt{2})^2}{4.10^{-21}C_{hold}} = 10^{64/10} \tag{3.9}$$

So a hold capacitor of $C_{hold} = 0.5$ pF results.

The switch transistor must be fast, minimum length is chosen. For a 1-GHz bandwidth a sample rate of 2 G/s is needed with a maximum track-time equal to half of the sampling period. For 60-dB 0.1 % settling is needed or $T_s/2 = 7\tau(e^{-7} = 0.09\,\%)$

$$\tau = R_{switch}C_{hold} = \frac{C_{hold}}{(W/L)\beta_\square(V_{DD} - V_T - V_{in,max})} < 35\,\text{ps}. \tag{3.10}$$

The resulting $W/L = 200$. The width of the switch will be 20 μm.

The size of the source follower transistor requires three checks:

the width must be sufficient for 1 mA current, the small-signal bandwidth must meet 1 GHz, and the slew rate should allow 0.4 V_{pp} at 1 GHz.

$$I_D = (W/2L)\beta_\square(V_{GS} - V_T)^2 = 1\,\text{mA}, \rightarrow W/L > 300$$

$$g_m = (W/L)\beta_{\square}(V_{GS} - V_T) > 2\pi f_{in,max} C_{load} \rightarrow W/L > 120$$
$$I_D > C_{load} V_{peak} 2\pi f_{in,max} = 0.8\,\text{mA} \tag{3.11}$$

$W/L = 30/0.1\,\mu\text{m}$. This hand calculation gives a first order estimate: distortion and the attenuation of the source follower have not yet been included.

3.4.3 Switched-Capacitor T&H Circuits

A switched-capacitor circuit circumvents the problem of a wide common mode range at the input of the buffer. Figure 3.21 shows a regular switched-capacitor circuit used for sampling and processing an input signal. In track mode the input signal $V_{in}(t)$ charges the capacitor C_1 via switches S_1 and S_2. Depending on the required range and specifications S_1 can be implemented with a bootstrap switch, while S_2 serves as a bottom-plate sampling switch. When the switches are opened the momentary value of the signal at the n-th sample moment $t = nT_s$ is stored as a charge on C_1: $C_1 V_{in}(nT_s)$. After closing the transfer switch S_3, the charge on C_1 is moved to C_2 and $V_{out}(t) = V_{in}(nT_s)C_1/C_2$. When the capacitors are equal and the opamp gain is high, this topology delivers an exact replica of the input signal at $t = nT_s$.

There are a few remarks to be made. Looking at the basic topology in hold mode, the opamp configuration is a 2× amplifier because its feedback factor is: $C_2/(C_1 + C_2) = 1/2$. The signal transfer has a 1× gain while undesired signals, such as contributions from input referred opamp noise, are multiplied by a factor 2. The basic transfer function, see the lower part of Fig. 3.21, shows a first order roll-off behavior with a pole at half of the unity-gain bandwidth. The opamp settling and the overall speed of the T&H are determined by this pole. Any offset at the opamp input is added to the signal.

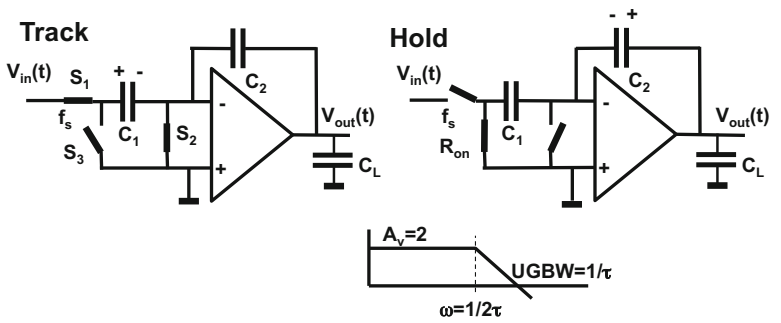


Fig. 3.21 A switched-capacitor circuit in track mode (*left*) and transfer (*right*)

In this topology the input is sampled when the track phase ends. Thermal energy allows electrons to move from input to the capacitor C_1 and back: here is a thermal noise source. When S_1 or S_2 opens the noise is sampled and held at the last value, resulting in a sampled noise contribution, whose long-term RMS value is: $v_{noise,rms} = \sqrt{kT/C_1}$. This frozen noise contribution is transferred into C_2 as if it were signal charge. Note that these noise samples are independent drawings from a statistical Gaussian amplitude distribution with characteristics: the mean equals $v_{mean} = 0$ and the standard deviation is: $v_{rms} = \sqrt{kT/C}$.

In hold mode S_3 becomes conductive: again electrons fluctuate between ground and C_1 through their thermal energy. This noise appears during the hold phase as time-continuous noise at the output of the opamp on top of the frozen noise from the initial kT/C_1 noise sample. When opening S_3 at the end of the hold phase, the path for electrons that move between the input and the capacitor C_1, is interrupted. The electrons will stay where they were. The noise is sampled and held, resulting in a second sampled noise contribution with the same long-term RMS value: $\sqrt{kT/C_1}$. This frozen noise contribution is also transferred into C_2 as if it were signal charge. This topology adds during one full cycle a noise sample with $v_{noise,rms} = \sqrt{2kT/C_1}$ to the signal sample.

3.4.4 Flip-Around T&H Circuit

The flip-around topology in Fig. 3.22 uses a single capacitor. The opamp is switched in a unity-gain mode during tracking mode, thereby creating a virtual ground node on the right side of the capacitor C_{hold}. When switches S_1, S_2 are conductive, the input voltage charges the capacitor. In hold mode the feedback switch S_3 is made conductive and the left side of the capacitor is connected to the output terminal. The operational amplifier is now fed back via the capacitor and the output voltage must be identical to the stored level of the input signal in order to keep the same virtual ground voltage between its input terminals.

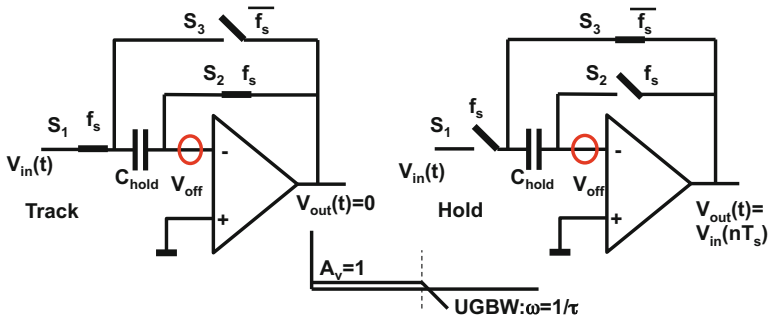


Fig. 3.22 A track-and-hold circuit with offset cancellation in track (*left*) and hold mode

In track mode the opamp is in unity-gain configuration. The resulting dominant pole is at the unity-gain bandwidth of the opamp. The tracking bandwidth is largely determined by the unity-gain of the opamp.

Settling in hold mode also happens with the opamp in unity-gain mode and is therefore faster than in the previous feedback topology. The amplification of input related noise is one.

When running on a low-supply voltage switch S_1 is bootstrapped. Switch S_2 is preferably a standard NMOS transistor, however, this choice depends on the value for the virtual ground. If the opamp has an NMOS input pair, the virtual ground may be of a level where S_2 must be chosen as a PMOS transistor or a bootstrapped NMOS transistor. Note that bottom-plate sampling can be implemented by allowing S_2 turn off before S_1 does. In medium speed circuits S_3 is often implemented as a transmission gate (PMOS parallel to NMOS). The flip-around topology has major advantages over other switched-capacitor topologies as in Fig. 3.21. The settling is faster, the noise is less and there is less area needed for capacitors.

Figure 3.23 shows a variant of the standard flip-around scheme. The operational amplifier is replaced by a transconductance stage. During tracking the capacitors are directly connected to the input signal and bias voltages. In feedback the capacitors are connected over the transconductance. This variant does not cancel the input offset of the transconductance, however, the kT/C noise contributes once and the noise of the transconductance during the track phase is not sampled in case this noise dominates.

3.4.5 *Offset and Noise in Flip-Around Circuits*

Any offset of the opamp or low-frequency noise will appear at the negative input terminal of the opamp and affect the virtual ground. The difference between the input signal and this offset is sampled and stored on the capacitor. During the hold phase the same offset is still there and will add to the capacitor voltage canceling

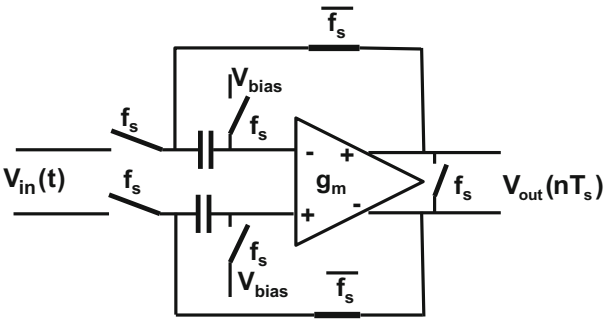


Fig. 3.23 A track-and-hold circuit based on a transconductance in differential design [41]

its effect. Consequently the output voltage is not affected by offset nor by low-frequency noise at the input of the opamp. The transfer of the offset source to the output signal is

$$V_{out,error} = V_{off}(z)(1 - z^{-0.5})$$

The exponent -0.5 assumes a 50 % duty cycle of the switching signal. The transfer in the frequency domain is written as:

$$H(f) = 2\sin(\pi f/2f_s) \tag{3.12}$$

This transfer function suppresses a DC offset voltage effectively. Though this scheme amplifies any (undesired) signal near $f_s/2$ by a factor of 2. So higher frequency noise components generated by the opamp are amplified and sampled into the signal (on top of the kT/C noise).

Similarly as in any T&H circuit the sampling process results in kT/C noise on the hold capacitor. During the initial track phase, time-continuous thermal noise is present on the capacitor. Next to this noise an input referred noise contribution of the opamp is present. With sufficient opamp gain, the noise generated inside the opamp is dominated by the noise of the transconductance g_m of the input pair. The feedback path returns this noise to the input terminal. As $1/g_m \gg R_{sw}$ the opamp contribution is relevant. $v_{gm,n}$ models this contribution in Fig. 3.24. The resulting output noise $v_{out,n}$ in track mode, experiences the (first-order) transfer function of the opamp. Combining the spectral density of the input referred noise for the transistors in the input pair and the unity-gain transfer gives:

$$v_{out,n}^2 = \frac{8kT}{3g_m}\int_{f=0}^{f=\infty}\frac{1}{1+(f/f_{UGBW})^2}df = \frac{8kT}{3g_m}\frac{f_{UGBW}\pi}{2} \tag{3.13}$$

leaving out any transconductance noise correction factor.[4]

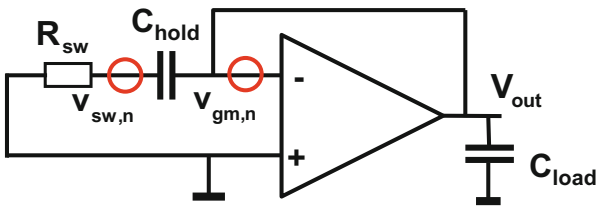


Fig. 3.24 The equivalent T&H circuit during track mode [42]

[4]The so-called noise-excess factor is disputed, although some correction for the shape of the inversion layer may be applicable.

The factor $f_{UGBW} \times \pi/2 \approx 1.57 f_{UGBW}$ accounts for the energy up to f_{UGBW} and the energy in the roll-off part of an ideal first order transfer curve.

The unity-gain frequency of the opamp is determined by the input transconductance of the input pair and the output load capacitance in case of a single stage opamp ($f_{UGBW} = g_m/2\pi(C_{load} + C_{hold})$, or by the transconductance and the Miller capacitance in case of a two-stage opamp ($f_{UGBW} = g_m/2\pi C_{Miller}$). The resulting noise is found by substituting these f_{UGBW} in Eq. 3.13:

$$v_{out,n}^2 = \frac{2kT/3}{(C_{load} + C_{hold}) \text{ or } C_{Miller}}$$

In this derivation the noise of the transconductance is modeled with a factor 8/3. For advanced processes this factor is probably closer to the resistive value of 4. In that case and with $C_{load} = 0$ the standard kT/C_{hold} re-appears. For a fully differential design the differential noise power doubles, while the signal power quadruples. So the signal-to-noise ratio increases by 3 dB by using two C_{hold} capacitors. The signal-to-noise ratio depends on the total amount of sampling capacitance, not on whether the circuit is single-ended or differential.

If the bandwidth of a two-stage opamp is comparable to the single stage version, the Miller capacitance should be of the same order of magnitude as the load capacitance. The load capacitance of the opamp certainly includes the hold capacitor. In total this means that the noise energy in the output during track mode is $kT/(C_{hold} + C_{load})$. This is a time-continuous noise limited by the opamp's bandwidth.

After the sample clock switches to hold mode, the noise in the output consists of two components: the sampled noise kT/C_{hold} which is a frozen value during the hold-period and time-continuous noise from the opamp kT/C_{load}:

$$v_{out,n}^2 = \frac{kT}{C_{hold}} + \frac{kT}{C_{load}}$$

During the hold mode this noise contribution on the output is time-continuous. In the succeeding processing this contribution is sampled and will be propagated as a charge packet.

A full cycle of switching activity in a switched-capacitor circuit will result in several separate and independent contributions of noise, whose energies are summed.

Example 3.4. Design a track-and-hold circuit based on Fig. 3.22. The process is 0.18 µm CMOS with a nominal power supply of 1.8 V, see Table 3.1. The output of the track-and-hold should be able to drive a 1 pF load. The sample rate is 10 Ms/s, and a 60-dB noise and accuracy are needed at the end of the sampling phase. Only NMOST switches can be used without bootstrapping. A generator with an impedance of 50 Ω generates a 0.4 $V_{peak-peak}$ signal and the DC input level can be chosen anywhere between ground and supply voltage.

Table 3.1 Transistor parameters for 0.18 µm CMOS as specified by ITRS [13] and various publications

	NMOS	PMOS
Threshold voltage V_T	0.39 V	−0.45 V
Current factor (3/3) $\beta_\square$	300 µA/V^2	80 µA/V^2
Output modulation λ	0.05 at $L_{minimum}$	0.05 at $L_{minimum}$
Output modulation λ	0.01 at $L = 1$ µm	0.05 at $L = 1$ µm
Back-bias factor γ	0.60 V$^{0.5}$	0.66 V$^{0.5}$

The NMOST has a threshold voltage of 0.4 V and a $\beta_\square = 350\,\mu A/V^2$. The PMOST has a threshold of 0.45 V and a $\beta_\square = 80\,\mu A/V^2$. Determine the size of the hold-capacitor, the switches, and type of opamp.

Solution. The architecture of Fig. 3.22 limits the input voltage range to a maximum of $V_{in,max} = V_{DD} - V_{T,NMOST} - V_{drive}$. If the drive voltage is chosen at 0.2 V, the input can vary between 0.8 and 1.2 V.

The total noise floor is calculated from the input root-mean-square signal level $0.4V/(2\sqrt{2})$) and the required SNR: 60 dB or 1000×, yielding 0.14 mV$_{rms}$. In the topology of Fig. 3.22 a kT/C_{hold} noise level is present due to sampling. Also during the hold mode (time-continuous) noise is produced of a similar power level. For calculating the necessary hold capacitor a safety factor of 2 is used: 0.14/2 mV$_{rms}$ results in $C_{hold} = 1$ pF.

At 10 MHz the track phase and the hold phase will last half a clock period or 50 ns. The settled value must be reached within 60 dB or $10^{-3} \approx e^{-7}$. Moreover, during the next track phase the capacitor must be discharged to a similar level. So settling time constant τ equals 50 ns/7$\approx$ 7 ns. This sets the requirement for the switches and the unity-gain bandwidth of the opamp. With $R_{tot}C_{hold} = 7$ ns, $R_{tot} = 7\,k\Omega$. Choosing a maximum switch resistance of $R_{sw} = 1.5\,k\Omega$ leaves some margin: $R_{sw} = L/W\beta_\square V_{drive}$ results in $W/L = 1.8/0.18\,\mu m$.

The unity-gain bandwidth of the opamp must exceed this speed level, so $2\pi f_{UGBW}\tau = 1$ and $f_{UGBW} > 20$ MHz. The settling level can only be achieved if the DC gain of the opamp is sufficient: $A_{DC} > 1000$. These numbers can be achieved with a Miller opamp.

A simulation of the circuit with these values will still reveal some distortion caused by the switches. Larger switches and more drive voltage are needed.

3.4.6 Amplifying T&H Circuit

A more complex track-and-hold uses the capacitors to multiply the signal. The circuit is shown in Fig. 3.25. During the track phase(left) the switch S_3 connects the operational amplifier in unity-gain feedback. Capacitors C_1 and C_2 are connected in parallel to the input signal. When the transition to the hold-phase occurs(right),

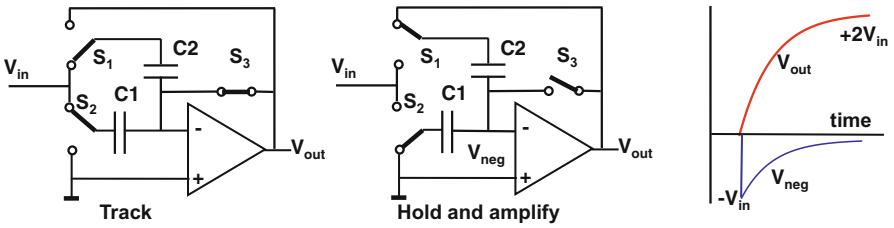


Fig. 3.25 A track-and-hold with multiplying capacitive feedback as is used for pipeline converters. *Right*: directly after switching to the amplify mode, there is a severe spike on the negative input terminal of the opamp

switch S_1 creates via C_2 the feedback path for the amplifier. S_2 switches capacitor C_1 to ground. The sampled charge on C_1 is transferred to C_2. The overall transfer of this circuit is

$$V_{out}((n+0.5)T_s) = \frac{(C_1+C_2)}{C_2} V_{in}(nT_s) \tag{3.14}$$

where $V_{out}((n+0.5)T_s)$ is the value at the output half a clock cycle after the corresponding input.

Figure 3.25 (right) shows the voltages on the negative input terminal of the opamp and the output voltage. Directly after switching, the voltage on the negative input terminal drops by the sampled value. As the same voltage is present over C_2, the output voltage will be zero (in absence of a load capacitor). Now the opamp starts reacting and will generate a positive output voltage. Ultimately this voltage will grow to $2V_{in}$ which compensates via the feedback factor $\beta = C_2/(C_1+C_2) = 0.5$ the initial step. The input spike can in some opamp topologies cause problems, and in extreme conditions (large input voltages and a near zero virtual ground level) the spike may open the source–substrate diode of the NMOS transistor implementing S_3. When signal charge is lost, the performance drops.

The initial sampling of the signal gives a noise power contribution of $v^2_{in,noise} = kT/(C_1+C_2)$. Or expressed as charge: $q^2_{in,noise} = kT(C_1+C_2)$.

In the amplification phase C_1 is connected to the ground level. The sampled signal and noise charge are stored in C_2. The opamp and the switches generate noise. Mostly the input transconductance is the main contributor. During the amplification phase this noise is filtered by the effective load of the amplifier. This additional contribution of noise has to be added to the sampled noise. A first estimate (ignoring the specifics of transconductances) is that the output noise is the thermal noise of the capacitive load: C_1 in series with C_2, yielding a power: $kT(C_1+C_2)/(C_1C_2)$. When divided by the power amplification $(C_1+C_2)^2/C_2^2$ this results in an input referred noise power of $kTC_2/(C_1C_2+C_2^2)$. The total input referred noise is therefore

$$v_{in,noise}^2 = \frac{kT}{C_1 + C_2} + \frac{kTC_2}{(C_1C_2 + C_2^2)} = \frac{kT}{C} \text{ in case } C_1 = C_2 = C$$

excluding the contributions of the switches and potential noise-excess factors. See for further discussion Example 8.10.

Although the exact specifications for the operational amplifier depend on the T&H architecture used, some general remarks apply.

As a rule of thumb[5]: the DC-gain of the operational amplifier should exceed the required performance numbers: for a 60 dB distortion performance the amplifier should be designed at 60–70 dB gain. In some analog-to-digital converter architectures, the performance depends on the accuracy of the charge transfer. A limited DC-gain will leave some charge on C_1 and impair this specification. Equation 3.8 sets a boundary condition.

As the opamp during the track phase is switched into a unity-gain configuration, the speed of the T&H is comparable to the unity-gain frequency. In the hold phase the topology of Fig. 3.25 is in an $A_v = (C_1 + C_2)/C_2$ amplification mode. The settling time constant of the topology equals $A_v/2\pi f_{UGBW}$. In a typical design the capacitors are equal ($A_v = 2$). If the sample rate equals half of the unity-gain bandwidth and the hold period is half of the sample period, there are π time constants to settle in the hold phase ($e^{-\pi} = 0.04$). If the unity-gain frequency is chosen at four times the sampling rate and various tricks are applied to stretch the hold period, an accuracy of 0.1 % is reached.

For high accuracy analog-to-digital converters the gain of the opamp must follow Eq. 3.8 to reduce incomplete charge transfer. The error voltage on the input of the opamp is equal to V_{out}/A_{DC}. In [43] the authors observe that reducing V_{out} will reduce the error, and a lower opamp gain can be tolerated. Figure 3.26 shows the basic concept. Both the track and amplify phases are identical to the standard solution. At the end of these modes capacitor C_3 is charged to the output voltage. In a third phase of operation, this capacitor is switched in series with the output terminal of the opamp. The opamp output voltage is reduced to virtually zero and consequently also the input related DC-offset voltage. The charge transfer is now

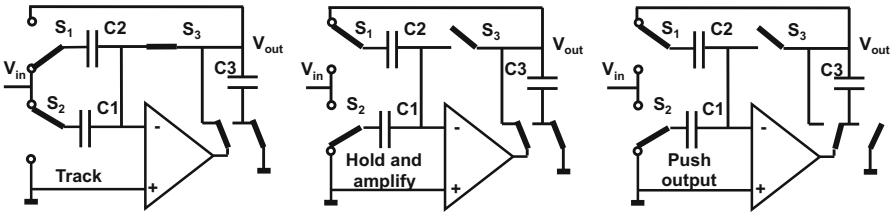


Fig. 3.26 An additional clock phase is used to reduce the requirements on opamp gain [43]

[5]Rule of thumb means here that this is a good level to start the discussion, various topologies give different results.

nearly perfect. This technique can also be applied inside a two-stage opamp, where it is inserted between the first and second stage, thereby allowing the output stage to deliver DC current. Also various multiplexing forms are conceivable. One penalty is in the timing; another in the signal transfer to the next stage.

3.4.7 Correlated Double Sampling

In most systems the track-and-hold circuit creates time-discrete samples from a time-continuous input. In some cases, the input is already time-discrete or is at least during a part of the sample period time-discrete. An example is the signal of an image sensor. During a clock period of such signal, various phases can be distinguished: a reset phase which sets the sensitive node to a predetermined value, a "zero-signal" phase, a period where the signal builds up and a signal-ready phase, see the V_C track in Fig. 3.27. The resetting of the sensitive sensor node introduces a number of unwanted components: noise from the reset transistor (kT/C), uncertainty of the charge splitting of the reset transistor, 1/f noise components, etc. The correlated double sampling technique eliminates a number of these detrimental effects by first sampling and storing a "zero-signal" level, that includes the before mentioned components. Then the signal builds up on the preset node. A second T&H samples the signal when the build-up is ready, and subtracts both. Figure 3.27 shows the basic functionality. In some designs, a single capacitor can be used for sampling both the "zero-signal" and the wanted signal. The transfer function for the unwanted components is

$$\begin{aligned} |H(s)| &= |1 - e^{-sT_d}| = |e^{-sT_d/2} 2\sin(sT_d/2)| \\ |H(\omega)| &= |2\sin(\omega T_d/2)| = |2\sin(\pi f T_d)| \end{aligned} \tag{3.15}$$

with T_d as the delay between the two sample moments. The main disadvantage also becomes clear from this transfer function: the noise near odd multiples of half of the sampling frequency is amplified.

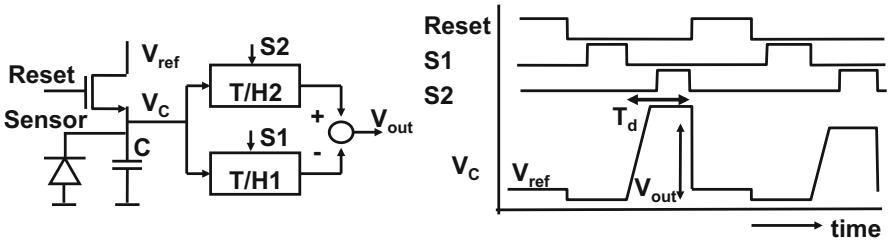


Fig. 3.27 The correlated double sampling technique

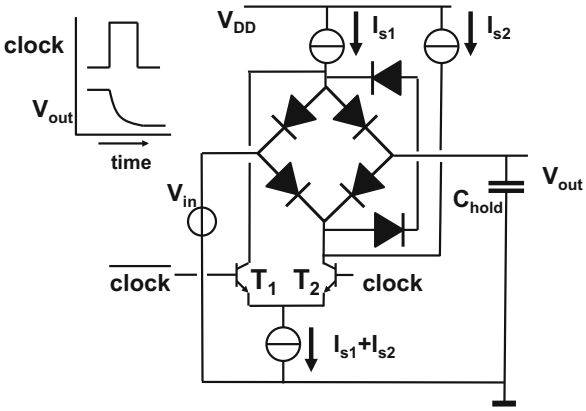


Fig. 3.28 A track-and-hold circuit with diode bridge [44]

3.4.8 Bipolar T&H Circuit

Bipolar design is still frequently used for analog realizations. Implementing a track-and-hold function with bipolar transistors requires a suitable switch topology. An often employed solution uses a diode bridge. Figure 3.28 shows a basic circuit.[6] When T_2 is conductive the current I_{s1} flows through the four-diode bridge. The typical relation between a diode current and its voltage is: $I = I_0 e^{qV/kT}$. After taking the derivative, the small-signal differential resistance is kT/qI. In this circuit the diodes carry each half of the current, so each shows a differential resistance of $2kT/qI_{s1}$. The path from input to output sees a resistance of $2kT/qI_{s1}$. With currents of 100 µA this corresponds to 500 Ω. In hold mode T_1 is active. The current I_{s1} is directly drained and the current I_{s2} flows via the two additional diodes, making sure that the bridge diodes are reverse biased and non-conductive.

The non-linear resistance in the conductive phase forces to use reduced signal swings. The leakage via the diodes is comparable to the leakage of diffusions in CMOS T&H circuits. With suitable diodes extremely fast switching is possible.

In contrast to complementary MOS technology (CMOS), commercial bipolar processes are not always equipped with two complementary device types with similar performance levels. The design style is therefore focussed on passing the signal only via high-performance npn-transistors. In [45] an often copied high-performance track-and-hold circuit in npn-only bipolar technology is presented. The circuit of which a single-ended version is shown in Fig. 3.29 is composed of a pre-stage, the actual switch and Capacitor, and an output buffer. T_1 and T_2 form an one-time input amplifier, buffering the signal towards the base of T_6. In the "Track" mode "T" is high and "H" is low. T_3 is switched off and the current of that branch passes via T_4 and biases T_6 as an emitter follower. The signal is applied to the hold capacitor C_H and T_7, T_8, and T_9 form an output buffer. In order to check the status

[6] The four-diode circuit is widely used for rectifying ac-signals under the name Grätz-bridge.

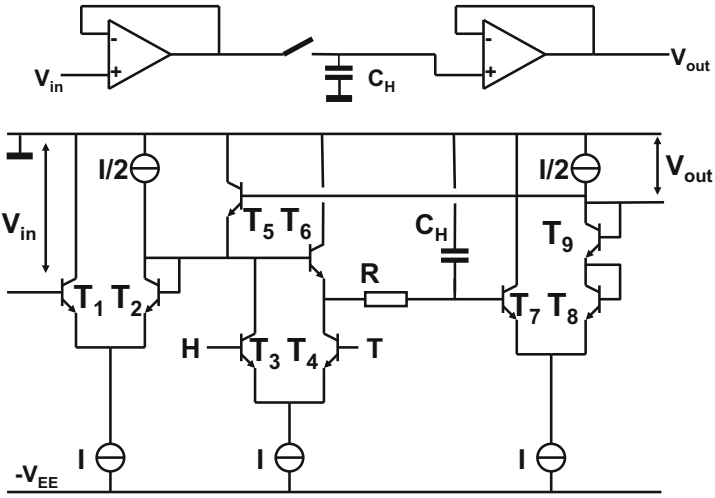


Fig. 3.29 A track-and-hold circuit in bipolar technology [45]

of T_5, the base-emitter voltages are counted: using the input or the base of T_1 as a starting point, the emitter of T_5 is at the same DC level as the input terminal. The voltage on the hold-capacitor is at V_{in} minus one V_{be} and the output at the collector of T_9 reaches the same DC-level as the input node. This voltage is applied to the base of T_5 which consequently has a zero base-emitter voltage and is not conducting.

Similar to the diode bridge switch, this design also uses current routing to perform the switch function. In the "Hold" mode the terminals "H" and "T" have reversed voltages. The current passes now via T_3 to T_5. Counting again the V_{be} voltages, starting from the hold capacitor node, it becomes clear that now T_6 has a zero base-emitter voltage and is not conducting. The sample-and-hold switch is non-conductive. If this circuit is designed as a pseudo-differential track-and-hold the droop-effect due to the base current of T_7 becomes a common effect. As long as the single-ended stages remain correctly biased the droop will only marginally affect the signal. Another improvement in the pseudo-differential circuit [45] is a pair of cross-connected compensation capacitors for reducing hold-mode feed-through.

Example 3.5. Can the hold capacitor in Fig. 3.29 that is now referenced to the positive power supply be connected to the negative supply rail?

Solution. With ideal supply voltages it does not matter whether the hold capacitor is referenced to the positive or negative supply. In practical circuits things are different. Due to the lack of good pnp-transistors it is often advantageous in bipolar design to use the positive rail as a signal reference. The power supply rejection for the negative power supply variations $-V_{EE}$ is now high because only current sources connect the circuit to the negative power supply. So all internal voltages are referenced to the positive power supply rail. Consequently the hold capacitor is

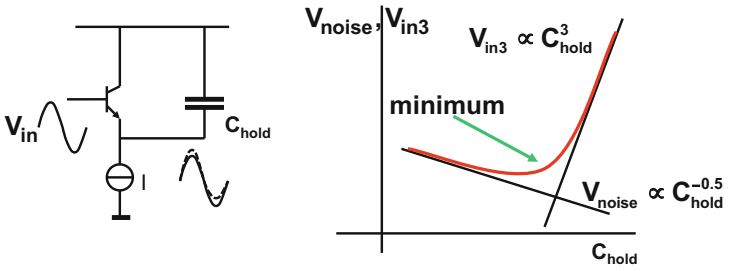


Fig. 3.30 This simple example shows that the hold capacitor cannot optimize the noise and distortion simultaneously

referenced to that rail as well, and the impact of variations on the negative power supply will be suppressed. These variations would directly add up to the held voltage on the capacitor if that capacitor is connected to the negative supply.

3.4.9 Distortion and Noise

The previous circuit example allows to illustrate a design choice that comes up in many track-and-hold circuits: the trade-off between distortion[7] and noise. The dependence of the noise voltage on the capacitance is known from kT/C Eq. 2.28. The distortion component can be calculated from the ideal circuit in Fig. 3.30. If the bipolar emitter follower circuit[8] copies the input signal to the capacitor and the current source provides an ideal and constant current, the capacitive current $i_C = j\omega C_{hold} V_a \sin(\omega t)$ will be taken away from the current in the bipolar transistor. This variable current will modulate the base-emitter voltage:

$$I - i_C = I_0 e^{q(V_{BE} - \Delta V_{BE})/kT}$$

Reversing the exponent to a logarithm and Taylor expansion for the first three terms gives:

$$\Delta V_{BE} = \frac{kT}{q}\left[\frac{i_C}{I} - \frac{1}{2}\left(\frac{i_C}{I}\right)^2 + \frac{1}{3}\left(\frac{i_C}{I}\right)^3\right]$$

If the input signal to the emitter follower is a perfect sine wave, the voltage that is rendered to the capacitor will differ ΔV_{BE} containing second and third order terms. The terms in the equation are a function of the signal current over the total

[7] See also [46] for an elementary discussion on distortion in analog circuits.

[8] A bipolar circuit is simple to analyze, of course the same holds for an MOS circuit.

current. This ratio, sometimes called "the modulation depth," determines the relative magnitude of the higher-order terms with respect to the fundamental. The second order and other even order harmonic terms can be eliminated by differential design, the third order and odd harmonics will, however, remain.

Substitution of the sine function leads to an estimate[9] of the fundamental and third order distortion component:

$$v_{C,1} = V_a - \frac{kT}{q}\left(\frac{V_a j\omega C_{hold}}{I}\right) \approx V_a$$
$$v_{C,3} = \frac{1}{12}\frac{kT}{q}\left(\frac{V_a \omega C_{hold}}{I}\right)^3 \tag{3.16}$$

More current or less signal swing, lower frequency or lower capacitance will reduce the relative impact of the third order term. On the other hand, less signal swing and a smaller hold capacitor lead to a lower signal-to-noise ratio. In the right-hand side of Fig. 3.30 the distortion term is compared to the noise contribution. For a given set of frequency, signal amplitude, and bias current parameters there will be an optimum capacitor value for a minimum combination of noise and distortion.

Example 3.6. If the third order signal component of a sample-and-hold is given by $v_3 = AC_{hold}^3$ and the noise signal level as $v_n = B/\sqrt{C_{hold}}$ what is the optimum C_{hold} for best SINAD performance?

Solution. Combining both components in the power domain gives:

$v_{tot} = \sqrt{(AC_{hold}^3)^2 + (B/\sqrt{C_{hold}})^2}$. Taking the derivative of v_{tot} with respect to C_{hold} and setting the result to 0 gives the minimum value:

$6A^2C_{hold}^7 = B^2$ where either the ratio $A/B = \sqrt{6C_{hold}^7}$ or the value of C_{hold} results.

Exercises

3.1. An FM signal of 100 kHz bandwidth at 10.7 MHz carrier is subsampled. Unfortunately a neighboring digital circuit running at 13.5 MHz will generate unwanted frequency components at 13.5 MHz and its second and third harmonic frequencies. Define a subsample frequency that maximizes the distance between the wanted and the spurious signals.

3.2. If the track-and-hold from Example 3.4 on page 79 is used to de-glitch a digital-to-analog converter and drive in a time-continuous mode an analog circuit, what should change?

[9] Contributions from higher-order terms are not included.

3.3. Calculate for the track-and-hold from Example 3.3 on page 73, the overall signal-to-noise ration and the pedestal step.

3.4. Calculate for the track-and-hold from Example 3.3 on page 73, the overall gain at DC and for a 1 GHz signal.

3.5. Propose for the track-and-hold from Example 3.3 on page 73, a bootstrap circuit, which parameters will change?

3.6. If the switch turns off at $t = T_1$ of the track-and-hold from Example 3.3 on page 73, sketch the output wave form taking into account that the output pole will influence the settling.

3.7. A simple T&H circuit as in Fig. 3.4 is designed in a 90-nm CMOS process with a switch transistor of dimensions $W/L = 5\,\mu\text{m}/0.1\,\mu\text{m}$. The hold capacitor is 100 fF and the input range is from 0 to 0.5 V. Give the overall amplification, the systematic offset, and the random offset of the capacitor voltage.

3.8. Discuss the impact of dummy switches in Fig. 3.6 on the random variation of the output offset voltage on the hold capacitor.

3.9. In an analog-to-digital converter the sampling speed is 250 Ms/s with an input range of $1.41\,V_{peak-peak}$. A performance of 90 dB is requested in a band from 30 to 40 MHz. The input sampling capacitor may contribute half of the noise energy. Calculate its size.

3.10. In sampling system the sample rate is 500 Ms/s with an input signal of range of $1\,V_{peak-peak}$ and a clock jitter of $1\,ps_{rms}$. A performance of 60 dB is required for all signals the Nyquist baseband. Calculate its size.

3.11. Consider in Fig. 3.6 the case where the compensation pulse is too early and the case where it is too late. Although both cases are unwanted, which will be less harmful to the performance?

3.12. A differential track-and-hold circuit shows a random offset. Design a switching sequence that performs the track-and-hold function and implements chopping to cancel the offset. Is there is practical solution for a single-sided track-and-hold circuit?

3.13. Use the technique in Fig. 3.26 between the first and second stage of a Miller opamp. What is the optimum position and what input error reduction can be achieved?

3.14. A signal is sampled on two parallel connected capacitors: C_1 and C_2, with $C_2 = 2C_1$. After sampling the capacitors are connected in series. Does the signal-to-noise ratio change between the parallel and series connection?

3.15. Modify Fig. 3.23 with a bottom-plate sampling technique.

3.16. A sine wave is sampled. There is a second order distortion present in the sampling chain. Is it possible to distinguish from the resulting output spectrum,

whether the distortion is affecting the original signal or appears after the sampling process?

3.17. Calculate in a 0.18 µm process the resistance of a complementary switch as in Fig. 3.10 with $W/L = 5/0.18\,\mu\text{m}$ at input voltages of 0, 0.5 V_{DD}, and V_{DD}. Use the data from Table 4.

Chapter 4
Quantization

4.1 Resolution

Quantization was, as many other techniques, first developed for telephony transmission purposes [47]. In this application the amplitude of a speech signal is represented by a number of pulses. This technique is called pulse code modulation (PCM). Today the term "PCM" is generalized and describes a quantized analog sample or its digital equivalent.

During quantization, the signal level at the sample moment is compared to a limited number of analog values, see Fig. 4.1. The amplitude of the continuous signal is rounded to the nearest level. In other words: the continuous amplitude is discretized. This process is called quantization.

An analog-to-digital conversion produces therefore rounding errors. The step from the analog to the digital domain is impaired with an error signal: the quantization error. The power associated with this quantization error is a fundamental limit to the quality of the process of analog-to-digital conversion.

Most amplitude discrete representations use a binary coding base as this format easily fits to the implementation of a converter by means of switches.[1] In a binary counting system one bit can have two values (0,1). A digital signal sample consists of a group of, e.g., 6, 10, or 16 bits. The number of bits needed to form a digital word representing an analog sample is called the word width or the resolution[2] N. In a binary coding scheme the number of levels is a power of the base number "2". The binary signal is limited to 2^N possible digital values and thus defines the number of levels to which an amplitude continuous signal is rounded.

[1] Other coding schemes will appear when discussing the implementation of complex converters, e.g., locally ternary (three-level) coding can be used in full differential implementations. In error correction schemes a base lower than 2 is applied.

[2] The IEEE has standardized a number of conversion terms in standards IEEE 1057 [48] and IEEE 1241 [49, 50].

M. Pelgrom, *Analog-to-Digital Conversion*, DOI 10.1007/978-3-319-44971-5_4

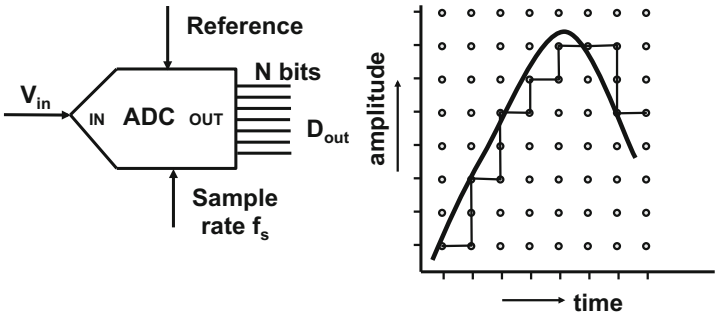


Fig. 4.1 An analog-to-digital converter rounds the amplitude of the continuous signal on the sampling moments to the closest value of the discrete amplitude scale

The accuracy of the conversion depends on the quality of the quantization levels. "Accuracy" and "resolution" are often confused. If the range between 0 and 0.8 V is theoretically quantized in 8 sections by 7 decision or trip levels: 0.1, 0.2 ... 0.7 V the resolution is $N = 3$ bit. In practice these levels are disturbed by offsets, gain errors, and random errors. If the decision levels are shifted to 0.04, 0.14, 0.24, ... 0.74 V, the relative error between the quantization levels is 0 %. The resolution is still perfect. But the absolute accuracy is shifted by 0.04 V with respect to an absolute external voltage level. In instruments like Volt-meters, the absolute accuracy and the resolution are crucial. In most communication systems only the resolution matters.

One possible signal representation in a binary system is the "straight binary representation"[3]:

$$B_s = \sum_{i=0}^{i=N-1} b_i 2^i = b_0 2^0 + b_1 2^1 + b_2 2^2 \ldots + b_{N-1} 2^{N-1} \tag{4.1}$$

The coefficient of the highest 2^{N-1} term b_{N-1} is called: "Most Significant Bit" or MSB, see Fig. 4.2. The coefficient of the lowest 2^0 term is b_0 and is called "Least Significant Bit" or LSB. If the LSB switches its value from 0 to 1, the next higher quantization level is selected. Therefore the step size between two successive quantization levels is also called "one LSB." The term LSB is in this book restricted to the numerical or digital domain. The physical equivalent of an LSB is written as A_{LSB} where A stands for voltages, currents, charges, or other physical quantities.

Now the relation between the analog and digital representation is

$$A_{LSB} = \frac{\text{physical range}}{2^N} \quad \Leftrightarrow \quad \frac{\text{full digital scale}}{2^N} = LSB \tag{4.2}$$

[3]Other binary code formats are discussed in Sect. 7.1.1.

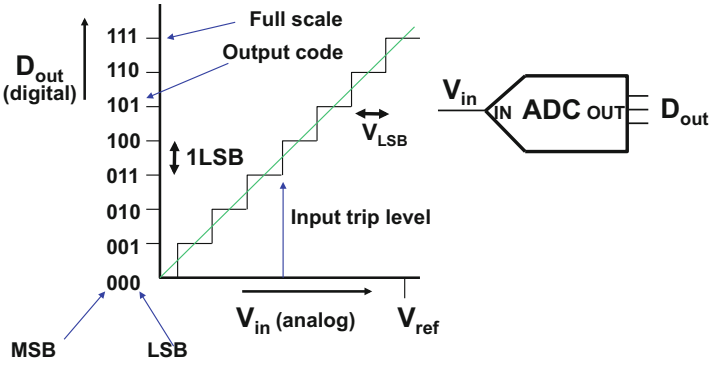


Fig. 4.2 Definition of analog-to-digital conversion parameters

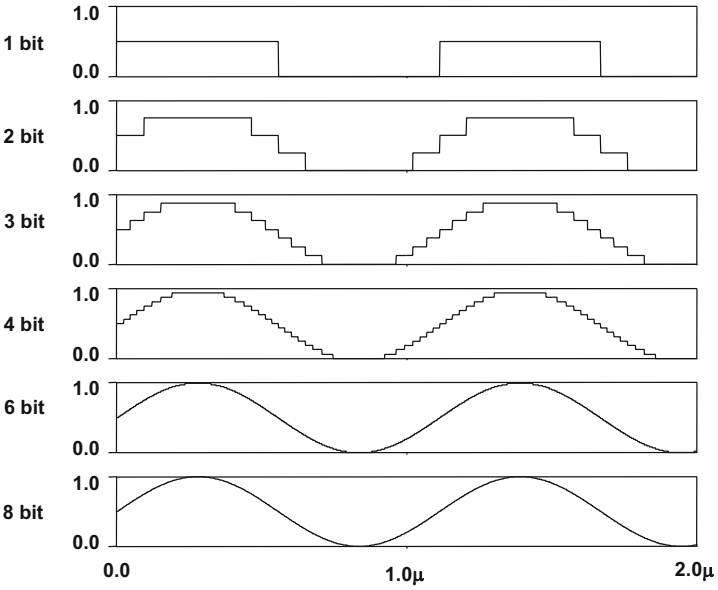


Fig. 4.3 Representation of a 900 kHz sine wave quantized with 1, 2, 3, 4, 6, and 8 bit resolution

where the physical range corresponds to the difference between the highest and lowest value of the expected analog signal. Of course there is no useful conversion available outside the range defined by the physical range. In order to warn users for such a situation, converters can be equipped with signals indicating an "overflow" or "underflow."

Figure 4.3 shows the quantization of a 900 kHz sine wave at an increasing resolution. A 1-bit resolution just signals the sign of the signal. At about 4 bits the signal shape becomes somewhat clear. Above 6 bit the human eye can hardly see the quantization steps in a plot.

4.2 Quantization Error

4.2.1 One-Bit Quantization

Quantization and sampling are mutually independent processes. The effect of sampling can be regarded as an independent pre- or post-processing of a quantized signal. In analog-to-digital conversion mostly the sampling of a time-continuous signal precedes the quantization. It is of course possible to start the signal processing with amplitude quantization: this small class of analog-to-digital converters is known under the name of level-crossing converters, see Sect. 8.9.1. The resolution of the conversion process after immediate amplitude quantization is carried by the timing of the resulting decision pulses. Which is theoretically equivalent, but practically an uneasy task.

Locking the digital signal value to a time frame and to an amplitude frame produces the quantization errors that limit the analog-to-digital converter's performance. The analysis of quantization errors is the subject of many mathematical treatises [51–53].[4] Figure 4.4 shows the spectrum of a 900 kHz signal quantized at 100 Ms/s with resolutions of $N = 1$, to $N = 8$.

In the case of $N = 1$ the analog-to-digital converter is in fact a simple one-level comparator and reshapes the input sine wave into a simple block wave. The quantization error equals the higher terms in the Fourier expansion of a block signal:

$$f(t) = \frac{4}{\pi}\sin(2\pi ft) + \frac{4}{3\pi}\sin(3 \times 2\pi ft) + \frac{4}{5\pi}\sin(5 \times 2\pi ft) + \ldots. \tag{4.3}$$

and the total power ratio between fundamental and harmonic components amounts to a theoretical value of

$$10\,{}^{10}\log\left(\frac{1}{\frac{\pi^2}{8} - 1}\right) = 6.31\,\text{dB}$$

4.2.2 2-6 Bit Quantization

When the signal is quantized with a resolution of more than one bit, the harmonic components will contain less power, as the approximation of the sine wave by the multilevel discrete signal will improve. Quantization is a distortion process with

[4] N. Blachman has mathematically analyzed many processes around quantization. His publications from 1960–1985 form a good starting point to dive deeper in this field, e.g., [52].

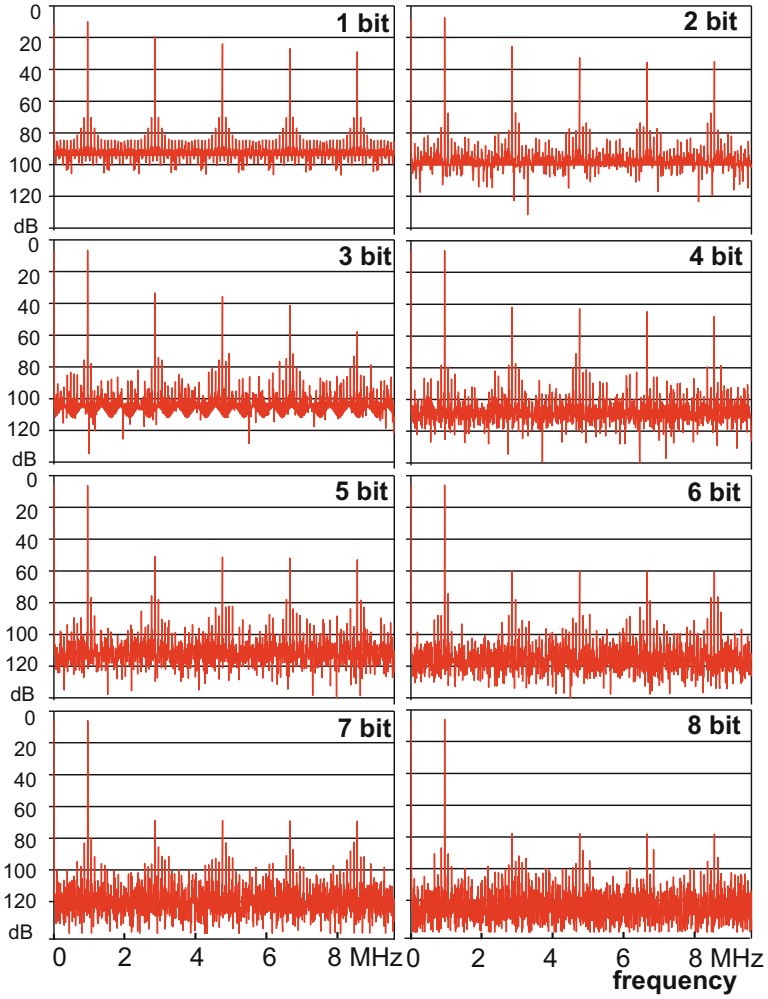


Fig. 4.4 The frequency spectrum of a 900 kHz signal quantized at 100 Ms/s with resolutions of $N = 1$ to $N = 8$. As the simulator used for generating this plot also samples the signal (10 µs), additional alias components of this process are visible, especially in the lower resolution spectra

significant higher-order components. Blachman derives a Fourier series for the p-th harmonic of the quantized signal $\hat{A}\sin(2\pi ft)$, where A_{LSB} is equal to 1 [52]. The resolution is increased by growing the sine wave amplitude $\hat{A} = 2^{N-1}$:

$$y(t) = \sum_{p=1,3,5,..}^{\infty} A_p \sin(2\pi pft)$$

$$A_p = \hat{A} \quad \text{for } p = 1,$$

$$A_p = \sum_{n=1}^{\infty} \frac{2}{n\pi} J_p(2n\pi\hat{A}) \quad \text{for } p = 3, 5, .., \tag{4.4}$$

A_p are the coefficients of the harmonic terms. J_p is a first order Bessel function. For large amplitudes of $\hat{A}$, the last equation can be approximated by:

$$A_p = \hat{A} \quad \text{for } p = 1,$$

$$A_p = (-1)^{(p-1)/2} \frac{h(\hat{A})}{\sqrt{\hat{A}}} \quad \text{for } p = 3, 5, ..,$$

In Blachman's theoretical analysis the fundamental component is $A_1 \propto 2^N$ and the odd distortion components $A_p \propto 2^{-N/2}$. This result means that ratio of the odd harmonics with respect to the fundamental is $2^{-1.5N}$ and reduces 9 dB per added resolution bit. Oude Alink et al. [53] finds a value closer to 8 dB per bit.

Figure 4.5 shows a simulation result of a 900 kHz fundamental frequency and some of its harmonic frequencies as a function of resolution. When the resolution is increased, a reduction of the third harmonic by some 8 dB per bit is seen. Also the amplitudes of higher-order harmonic components reduce in amplitude. Their curves are less regular as some harmonics interfere with aliased components in this simulation.

Even a perfect quantization will not nullify the lower harmonics.

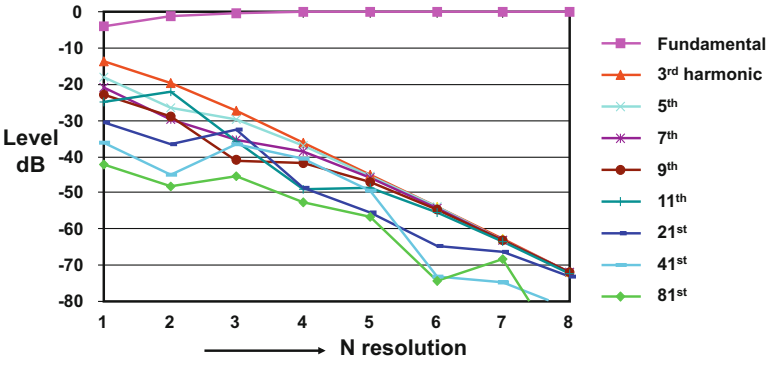


Fig. 4.5 The frequency components of a 900 kHz signal quantized at 100 Ms/s with resolutions of $N = 1$ to $N = 8$. The fundamental and the harmonic at $3, 5, 7, 9, 11, 21, 41$, and 81 times the fundamental frequency are shown as a function of the quantization resolution. A decay of 8–9 dB/bit is visible

4.2.3 7-Bit and Higher Quantization

The quantization error is a non-linear phenomena and an approximation can be used for quantization levels higher than 6 bit quantization. The power contained in the error signal plays a dominant role in this analysis of analog-to-digital converters. In an analog-to-digital converter of low-resolution the quantization error strongly depends on the input signal and will show up as distortion, as is shown in Fig. 4.4. However for a signal that passes a lot of quantization levels with a frequency that is not linked to the sample rate, the multitude of distortion products that mix with all multiples of the sample rate allow a statistical approximation of the error signal. The deterministic error signal after quantization is approximated as white noise in the band from 0 to $f_s/2$ and mirrored to the higher bands. Some authors call this quantization power: "noise," however, this power shows only in a limited sense noise properties, see Table 4.1.

This first order approximation for quantization of the signal supposes that the time-continuous signal has a uniform probability density for the occurrence of the signal within the range of conversion. This assumption does not take any specific signal properties into account, neither will the result be sensitive to small changes in the properties of the signal. Figure 4.6 shows the error signal that is generated while the input signal of an analog-to-digital converter steadily increases. The error signal is at its extremes at the trip levels, and varies linearly from $+0.5A_{LSB}$ to $-0.5A_{LSB}$. Optimum quantization or the lowest error power is reached if the quantization levels are uniformly spaced over the full amplitude scale [54, 55].

The probability density of the error signal between $+0.5A_{LSB}$ and $-0.5A_{LSB}$ is assumed constant and uniformly distributed. The power in that range is determined by calculating the estimation value for the variance, or the area of the triangles in Fig. 4.6. The equivalent power of the quantization is found from averaging the

Table 4.1 A comparison of thermal noise and quantization power

Thermal noise	Quantization power
Physical model describing random motion of electrons with thermal energy	Mathematical model for modulated distortion products
White noise density in frequency domain	White noise density in frequency domain
Temperature dependent	No physics involved
Component value dependent	Resolution dependent
Requires signal power to reduce noise	Requires more resolution to reduce errors
Amplitude is Gaussian distributed $\mu = 0$, $\sigma = v_{rms}$	Amplitude is uniformly distributed $[-A_{LSB}/2, A_{LSB}/2]$
$v_{rms} = \sqrt{4kTR\Delta BW} = 4\,\text{nV}$ $@R = 1\,\text{k}\Omega, \Delta BW = 1\,\text{Hz}$	
$v_{rms} = \sqrt{kT/C} = 0.22\,\text{mV}$ $@C = 1pF, BW = 0..f_s/2$	$A_{rms} = \frac{A_{LSB}}{\sqrt{12}}$ $BW = 0..f_s/2$

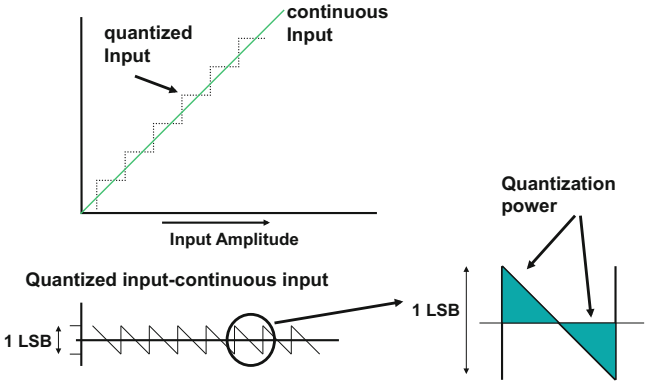


Fig. 4.6 As the analog-to-digital converter rounds the signal to the closest level, a *triangular shaped* error signal will arise. This saw-tooth signal is the fundamental quantization error

amplitude power. The quantization power[5] now equals [51]

$$Q^2 = \frac{1}{A_{LSB}} \int_{\varepsilon=-0.5A_{LSB}}^{\varepsilon=0.5A_{LSB}} A_{error}^2(\varepsilon)d\varepsilon = \frac{1}{A_{LSB}} \int_{\varepsilon=-0.5A_{LSB}}^{\varepsilon=0.5A_{LSB}} \varepsilon^2 d\varepsilon = \frac{A_{LSB}^2}{12} \tag{4.5}$$

Despite the above mentioned constraints, such as uniform distribution and constantly spaced quantization levels, this formula is sufficiently accurate for many applications.

The term A_{LSB} refers to the physical size of the least significant bit. Referred to the full-scale:

$$\text{Quantization power} = \frac{A_{LSB}^2}{12} = \left(\frac{\text{full-scale}}{2^N\sqrt{12}}\right)^2 \tag{4.6}$$

where "full-scale" is the analog range of the conversion and N stands for the resolution.

Power or energy forms the basic dimensions to express signals, noise, and other artifacts. As long as various signal elements are mutually independent, their power or energies can be added together. This is a consequence of the first law of

[5]Note that in a formal sense just voltage-squared is calculated, which lacks the impedance level and the time span needed to reach the dimension of power: Watt or $V^2/\Omega s$. In quantization theory "voltage-squared" power is only used to compare to another "voltage-squared" power, assuming that both relate to the same impedance level and the same time span. This quantization error becomes visible to an engineer (mostly) as part of a power spectrum. For that reason this book prefers the term quantization power instead of energy.

thermodynamics, stating the conservation of energy principle. Signals in the voltage or current domain become interesting when considering the circuit implementation.

Example 4.1. An ideal 10-bit analog-to-digital converter with a full-scale range of 1.024 V is used to process a 15 $mV_{peak-peak}$ sine wave signal. What model for the quantization error should be applied?

Solution. The 10-bit resolution of the converter seems to point to the quantization noise approximation (as $N > 6$). However, a closer look reveals that the LEB size is 1 mV, so the signal spans just 15 quantization levels, or 4 bit. Obviously it is better to interpret the resulting errors as the products of distortion. Inspection of Fig. 4.5 shows that distortion products at a −40 dB level can be expected.

4.3 Signal-to-Quantization Noise

At a resolution of $N = 7$ or higher the spectrum of an error signal is mostly flat enough for an approximation as "white noise." Yet the power in the lower harmonic distortion components can be relevant. In specific (communication) systems still special attention must be paid to the distortion in the quantization error. When the assumptions for the white noise approximation hold, the "noise" due to quantization often dominates over other noise sources up to 14 bit resolution.

The quantization errors of the analog-to-digital converter are considered to be mutually independent. The resulting white noise power stretches in the frequency domain from $f = 0$ and $f = f_s/2$ and is modulated around the multiples of the sample rate. The quantization error spectral power density (power per Hz) is therefore equal to the total quantization power divided by the bandwidth

$$S_Q(f) = \frac{A_{LSB}^2/12}{f_s/2} = \frac{A_{LSB}^2}{6f_s} \tag{4.7}$$

In many systems the signal bandwidth BW is fixed. The accumulated quantization error power is $S_Q(f) \times BW$.

4.3.1 SN_QR

Many system definitions use sine wave related specifications. In a practical measurement an almost perfect sine wave can be generated with the help of high-quality filters. Therefore sinusoidal signals are preferably used to characterize the performance of analog-to-digital converters. The signal-to-quantization error ratio is consequently expressed as the ratio between the maximum sinusoidal signal power that the analog-to-digital converter can handle and the quantization power. In fact this comparison violates the above assumption of a uniformly distributed signal over

the conversion range. This violation occurs especially where the sine wave is at its peak values. Its consequences are mostly acceptable.

The signal-to-quantization-noise ratio (SN_QR) compares the quantization error power to the power of a full sine wave in a similar fashion to characterizing thermal noise:

$$SN_QR = 10\,^{10}\log\left(\frac{\text{Signal power}}{\text{Quantization error power}}\right) = 20\,^{10}\log\left(\frac{V_{signal,rms}}{V_{Q,rms}}\right) \tag{4.8}$$

Substituting the signal power and quantization error power gives the signal-to-quantization-noise ratio:

$$\text{Signal power} = \frac{1}{T}\int_{t=0}^{t=T} \hat{A}^2 \sin^2(\omega t)dt = \frac{\hat{A}^2}{2} = \frac{2^{2N}A_{LSB}^2}{8}$$

$$\text{Quantization power} = \frac{A_{LSB}^2}{12}$$

$$SN_QR = \frac{\text{Signal power}}{\text{Quantization power}} = \frac{3}{2}2^{2N}$$

$$SN_QR \text{ in dB} = 10\,^{10}\log\frac{3}{2}2^{2N} = 1.76 + N \times 6.02\ dB \tag{4.9}$$

This last formula is an approximation that is often used for designing an analog-to-digital converter:

The maximum signal-to-quantization-noise ratio (SN_QR) represented by a digital word of N bits in a bandwidth of $f_s/2$ is 1.76+6.02xN dB.

All simple descriptions have their limits, so the question arises: when is the white noise model a correct approximation for a phenomena that in fact is a complex distortion and down-sampled spectrum. Table 4.2 compares the simulated signal-to-quantization power ratio to the simple approximation of $6.02N+1.76$ dB. For $N = 1$ the simulated value corresponds perfectly with the mathematical analysis. With

Table 4.2 Simulated signal-to-quantization power compared to the approximation formula

Resolution	Simulated SN_QR	$6.02N + 1.76$ dB	Theory
1	6.31 dB	7.78 dB	6.31 dB
2	13.30 dB	13.80 dB	
3	19.52 dB	19.82 dB	
4	25.60 dB	25.84 dB	
5	31.66 dB	31.86 dB	
6	37.71 dB	37.88 dB	
7	43.76 dB	43.90 dB	
8	49.80 dB	49.92 dB	

increasing N the approximation gains accuracy, although a minor overestimation of the quantization $\mathrm{SN}_Q\mathrm{R}$ remains. A cause for this overestimation was earlier identified: the uniformity of the signal over the entire range is assumed, which is not the case for a sine wave near the top and bottom values.

In video applications the top–top signal value of the video content is used as a reference. This description results normally in a 10 dB higher value for the SNR than in the sinusoidal case.

Note that when the signal amplitude spans 3% of the full range of a ten-bit analog-to-digital converter, this conversion functionally is equivalent to a five-bit analog-to-digital converter.

Example 4.2. What signal-to-quantization-noise ratio can be reached when the digital signal is transported over an 8-bit digital bus?

Solution. The best performance for full bandwidth is limited to $\mathrm{SN}_Q\mathrm{R} = 8{\times}6.02 + 1.76 = 50\,\mathrm{dB}$.

4.3.2 Related Definitions

In signal processing ratios between various quantities (signals, noise, and distortion) are specified as power ratios, that compare the power of noise, jitter, and distortion to the power of the maximum sine wave $v_{signal}(t) = \hat{A}\sin(\omega t)$.

The signal-to-noise ratio (SNR) is specified as:

$$\mathrm{SNR} = \frac{\mathrm{P_{signal}}}{\mathrm{P_{noise}}} \text{ or } \mathrm{SNR} = 10^{10}\log\left(\frac{\mathrm{P_{signal}}}{\mathrm{P_{noise}}}\right) \tag{4.10}$$

While the total harmonic distortion (THD) uses (somewhat confusing) the inverse ratio [46]:

$$\mathrm{THD} = \frac{P_{\mathrm{distortion}}}{P_{\mathrm{fundamental\ harmonic}}}$$

In audio engineering the THD is expressed in %. As these ratios can amount many orders of magnitude, a logarithmic notation often replaces the exponential notation[6]:

$$\mathrm{THD} = 10^{10}\log\left(\frac{P_{\mathrm{distortion}}}{P_{\mathrm{fundamental}}}\right) = 20^{10}\log\left(\frac{V_{\mathrm{distortion}}}{V_{\mathrm{fundamental}}}\right) \tag{4.11}$$

The signal-to-noise-and-distortion (SINAD or SNDR) stands for the ratio of the signal power to all the unwanted components: quantization errors, thermal noise, distortion, etc.:

[6] 43.8 dB is a short hand for 4.167×10^{-5} power ratio. Use the exponential notation in complex calculations.

$$\text{SINAD} = 10\,{}^{10}\log\left(\frac{\text{Power of 1st harmonic}}{\text{Power of all unwanted components}}\right) \tag{4.12}$$

The spurious free dynamic range (SFDR) is the distance between the signal and the largest single unwanted component, the spurious signal.

The dynamic range (DR) represents the ratio between the full-scale input signal and the noise floor at a small-signal input. The difference between DR and SNR is present in, e.g., range switching configurations and in sample systems with jitter. Here the dynamic range can be higher than the signal-to-noise ratio, see Fig. 2.27.

In order to characterize the converter in a simple manner, the effective number of bits (ENOB) is calculated by reversing equation 4.9:

$$\text{ENOB} = \frac{\text{SINAD} - 1.76}{6.02} \tag{4.13}$$

The ENOB allows an easy comparison of the real performance of a converter. Suppose a nominal 8 bit resolution is in the data sheet. If the measurements show only 6.8 ENOB, the converter loses a lot of performance. At 8-bit level no more than 0.5 ENOB should be lost. At 12-bit a loss of 1 ENOB is tolerable.

ENOB is a resolution-like parameter and 2^{ENOB} is a signal ratio, not a power ratio.

Example 4.3. What signal-to-noise ratio can be reached by a 10-bit 50 Ms/s analog-to-digital converter in the bandwidth from DC to half of the sampling rate? And in a bandwidth between 2.0 and 2.5 MHz?

Solution. A converter of 10 bit resolution can reach a signal-to-noise ratio in the full conversion band of 0–25 MHz of $\text{SN}_Q\text{R} = 1.76 + 6.02 \times N = 62\,\text{dB}$.

In a bandwidth limited to 0.5 MHz, only the proportional fraction of the total quantization power will appear: 0.5/25. In power ratios this factor corresponds to ${}^{10}\log 50 = 17\,\text{dB}$. Therefore the signal-to-quantization error in a limited bandwidth is $\text{SN}_Q\text{R}_Q = 62 + 17 = 79\,\text{dB}$.

Example 4.4. An ideal 8-bit quantizer samples at 60 Ms/s. In the following digital circuit a filter section limits the band of interest to 1 MHz, what is the expected full signal-to-quantization power ratio in this band. What combinations of resolution and sampling speeds are possible for digital representation of this signal next to an 8-bit sample at 60 Ms/s.

Solution. The signal-to-quantization-noise of an ideal 8-bit quantizer is given by: $\text{SN}_Q\text{R}=1.76 + 6.02 \times N = 49.9\,\text{dB}$. The quantization power is spread out evenly over the band from 0 to $f_s/2$ and then mirrored to higher bands. With $f_s = 60$ Ms/s the quantization power is contained in a 30 MHz wide band. In a $\text{BW} = 1$ MHz band there will a factor 30 less power or in dB: $10^{10}log(30) = 14.8\,\text{dB}$. That brings the SN_QR in the 1 MHz band to 64.7 dB.

Table 4.3 Combinations of resolutions and sample rates resulting in the same SN_QR (Ms/s)

N	f_s
5	3840
6	960
7	240
8	60
9	15
10	3.75
11	2

The generalized formula linking the sample rate to the SN_QR in a bandwidth BW is

$$SN_QR(BW) = 1.76 + 6.02 \times N + 10^{10} \log(\frac{f_s/2}{BW})$$

Keeping the SN_QR fixed to 64.7 results in the values of Table 4.3. Note that for a resolution of 11 bit or higher the sample rate is fixed by the lower boundary of the Nyquist criterium.

4.3.3 Non-uniform Quantization

In many systems the relevance of a small variation on the amplitude is different for various levels of the amplitude. Video signals show most detail close to the black level, while in full-white signals only little information is present. The reason is in the logarithmic signal processing of the human perceptory organs. Speech signals are symmetrically distributed around the zero value and need much finer quantization around the mid-point than in the extreme excursions. A similar argument holds for OFDM modulated digital communication signals. These signals consist of a large number of individually modulated carriers. On average the power of these carriers will add up in a root-mean-square sense: $v_{rms} = \sqrt{v_1^2 + v_2^2 + \ldots}$, however, every now and then a time moment may occur where a significant number of carriers are in phase and their amplitudes add up: $v_{top} = v_1 + v_2 + \ldots$. The ratio between v_{top} and v_{rms} is called the crest factor or the peak-to-average ratio. Sine waves have a crest factor of $\sqrt{2}$ or 3 dB. In practical communication systems crest factors in excess of 10 dB will occur, requiring more than a full bit of extra resolution.

In communication literature various attempts have been made to come up with specific non-uniform quantization schemes, to improve the conversion quality at lower hardware costs, e.g., for a normal amplitude distribution [55]. However in practice this requires designing a specific analog-to-digital converter for each signal type. A more common approach is to precede a uniform quantizer with

an appropriate compression circuit. A starting point is the μ-law compression of function $f(t)$ in a compressed $g(t)$ function:

$$g(t) = \text{sign}(f(t)) \frac{\ln(1 + \mu|f(t)|)}{\ln(1 + \mu)} \tag{4.14}$$

where μ determines the amount of compression. It is obvious that after compression and analog-to-digital conversion a de-compression step is required with the exact inverse function. Any intermediate signal processing must take the complex structure of the compressed signal into account. Other solutions implement floating-point architectures such as variable gain amplifiers or two parallel analog-to-digital converters, see Sect. 8.9.5.

4.3.4 Dither

Variations in a signal that occur between two trip levels will be lost after rounding to the closest representation level. This quantization distortion can be annoying in low-resolution conversion systems. A helper signal or "dither" can be added to the original signal [56, 57], see Fig. 4.7. This dither signal has typically the magnitude of $1\,A_{LSB}$ and consists of a well-chosen (random) signal.[7] The dither should of course not interfere with the wanted signals. Options for implementation are: at high sample rates out-of-band dither can be applied, the dither is a known signal that can be extracted or subtracted in the digital domain, or the dither is a low-amplitude random signal.

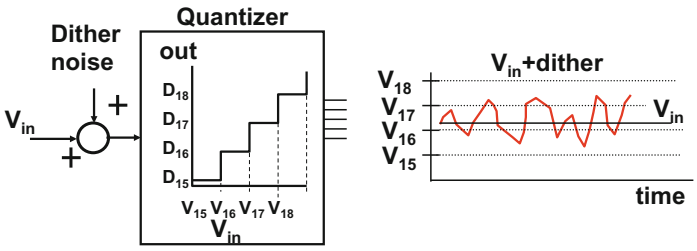


Fig. 4.7 The addition of a random signal allows to determine the value of a DC-signal at greater accuracy than the quantization process allows

[7] There has been an extensive search for optimum dither signals in the 1960–1970s. After that era the interest for dither has reduced. The concept, however, still provides valuable insight, e.g., sigma-delta converters can be understood as low-resolution converters that generate their own dither.

In Fig. 4.7 a high-speed quantizer with low-resolution converts a slow changing signal. A high-frequency dither addition pushes the signal over and under the trip levels. The probability of crossing the trip levels depends on the distance of the original signal level to the trip levels. Additional signal processing like averaging can now lead to resolution improvement for low-frequency signals (compare also Sect. 10.1). In spectrum terms: the dither signal turns the quantization distortion into high-frequency random noise.

Conceptually the most simple form of dither is thermal noise which is, however, not so easily controlled. An alternative is uniform distributed noise. When an input signal V_{in} is quantized to $B_{integer}$ with a quantization loss fraction $-0.5 \leq Q_{fraction} < 0.5$, the following equivalence holds:

$$V_{in} = (B_{integer} + Q_{fraction}) \times V_{LSB} \tag{4.15}$$

The addition of the uniform dither ranging from $-0.5\,\mathrm{V}_{LSB}$ to $0.5\,\mathrm{V}_{LSB}$ activates randomly the quantization trip levels around $B_{integer}$. For a large number of M samples, the amount of hits on these levels is M_{B-1} and M_{B+1} and related to as $Q_{fraction} \propto (M_{B+1} - M_{B-1})$. The average digital output over M samples is now:

$$\begin{aligned} &\frac{(M - M_{B+1} - M_{B-1})B_{integer} + M_{B+1}(B_{integer} + 1) - M_{B-1}(B_{integer} - 1)}{M} \\ &= B_{integer} + \frac{(M_{B+1} - M_{B-1})}{M} \approx B_{integer} + Q_{fraction}\big|_{M\to\infty} \end{aligned} \tag{4.16}$$

This example of uniformly distributed dither with an amplitude of V_{LSB} may seem constructed and difficult to realize in practice, however, note that the noise shaper in Sect. 10.2 uses its own delayed quantization error as a uniformly distributed dither signal.

In [56] the theoretical background of the use of dither in quantization is explored. The authors show that dither can be used to de-correlate errors in conversion. Distortion products are converted into white noise-like phenomena. But the total quantization power itself is not reduced. This approach is often applied in RF conversion systems where tones cannot be tolerated. Next to dithering in the time-domain, also dither in the spatial domain can be used. If multiple channels with essentially the same structure are applied in parallel, dither in the form of random offsets is applied to each channel. This form of dithering allows to randomize the quantization errors [58]. With multiple channels some reduction in quantization power can be achieved. This is easily understood, e.g., a second channel with a $0.5\,\mathrm{V}_{lsb}$ offset allows to increase the resolution by 1 bit.

4.4 Linearity

An ideal analog-to-digital converter will only show the imperfections due to the quantization process. In practice there will be many additional errors, depending on the method of conversion, the available technology and circuit, the required resolution, the signal and sampling frequency, and the skills of the designer.

4.4.1 Integral Linearity

Figure 4.8 shows a staircase formed by the analog values where the digital code trips from code i to $i+1$. The "trip levels" or "decision levels" of an ideal converter are given by $i \times A_{LSB}$, where A_{LSB} is the LSB step. The curve $i \times A_{LSB}, \forall i = 0, \ldots, (2^N - 1)$, forms the ideal conversion curve for a resolution N.

The integral non-linearity (INL) plot shows the deviation of a practical conversion curve given by $A(i), \forall i = 0, \ldots, (2^N - 1)$, from the ideal conversion function.

$$\mathrm{INL}(i) = \frac{A(i) - i \times A_{LSB}}{A_{LSB}}, \ \forall i = 0..(2^N - 1) \tag{4.17}$$

In an analog-to-digital converter $A(i)$ stands for the trip level of the input parameter. In a digital-to-analog converter, $A(i)$ is the output value for the corresponding digital code.

In Fig. 4.9 the error is plotted in LSB units, while the horizontal axis shows the ideal digital code. This normalizes the plot and allows an easy analysis and interpretation. The integral linearity plot given by INL(i) reveals issues with the

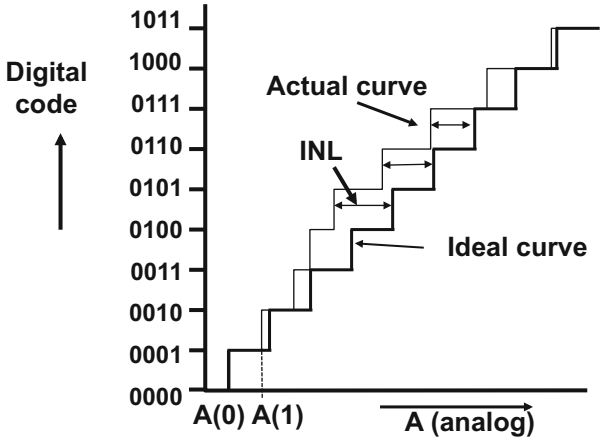


Fig. 4.8 Definition of the integral linearity error

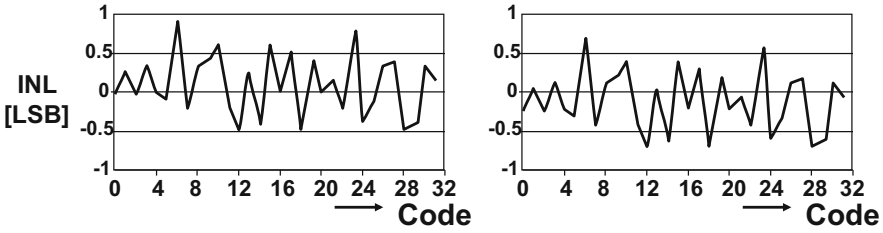


Fig. 4.9 A typical integral non-linearity plot for a 5-bit analog-to-digital converter. *Right*: the *curve* is shifted and rotated to get the best-fitting *straight line*

converter and also discloses some information on the internal structure of the converter, see Sect. 4.5

Often the INL is given as a curve, however, INL can be reduced to two numbers corresponding to the maximum positive and negative deviation over the entire range:

$$\text{INL} = \max, \min\left(\frac{A(i) - i \times A_{LSB}}{A_{LSB}}\right), \ \forall i = 0..(2^N - 1) \tag{4.18}$$

The maximum and minimum number for the INL of the converter in Fig. 4.9 is therefore +0.9LSB/−0.5LSB. The above definition implies that the conversion process starts at an input signal at "0" level and ends at the full-scale value. This absolute accuracy requirement is important in some industrial and measurement systems. In many other systems offsets in the absolute value are acceptable, e.g., in accoupled circuits. Deviations in the slope of the transfer curve can also be handled by many systems and result in a negligible amplification error.

In those cases a more loose definition of the integral linearity is sufficient: the deviations in INL(i) are then measured against the best-fitting straight line. In the example of Fig. 4.9 this will mean a shift of roughly 0.2 LSB, which results in an INL specification of +0.7/−0.7. See also Fig. 4.14 (left, middle). The maximum INL specification is now a lower number, although nothing has changed in the performance.[8]

The integral linearity is directly related to the harmonic distortion, see also Sect. 4.5. The specific shape of the transfer curve of a converter as it is given by INL(i) will determine the magnitude of the individual harmonic components. The power of the errors is reflected in the total harmonic distortion (THD). Equation 4.12 defines the THD as the power ratio of the harmonics of a signal and the fundamental tone. Usually for analog-to-digital converters the first 5 or 10 harmonics are counted as THD, while higher-order components and folded products are counted as SINAD contributions.

[8] A manipulation like this is aversely coined: "specmanship."

In communication and audio systems the overall equipment performance depends on the THD and therefore the INL curve is very relevant in converters aimed at these markets.

4.4.2 *Differential Linearity*

Next to the integral linearity the differential linearity is important in characterizing the DC-transfer curve of analog-to-digital and digital-to-analog converters. The differential non-linearity (DNL) is the deviation of each step with respect to the ideal LSB size. The mathematical formulation is

$$\mathrm{DNL} = \frac{A(i+1) - A(i)}{A_{LSB}} - 1, \ \forall i = 0..(2^N - 2) \tag{4.19}$$

or as a single maximum number:

$$\mathrm{DNL} = \max \left| \frac{A(i+1) - A(i)}{A_{LSB}} - 1 \right|, \ \forall i = 0..(2^N - 2) \tag{4.20}$$

Figure 4.10 shows in the lower part two limited DNL errors. For clarity small bars are added to show the size of "1 A_{LSB}." Higher up in the curve two extreme situations of DNL errors are illustrated. In certain constructions of converters, such as binary coded analog-to-digital converters, an increasing input signal may result in a step towards a lower digital output code: the converter is non-monotonic.[9] This behavior can result in catastrophic problems if this converter is applied in a control loop.

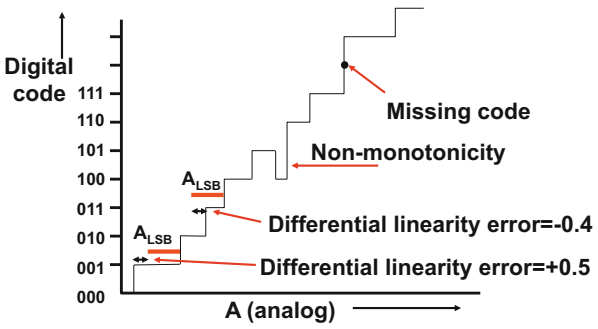


Fig. 4.10 Definition of the differential linearity

[9]Non-English speakers often confuse monotonic with monotonous which is synonymous to boring, dull, and uninteresting.

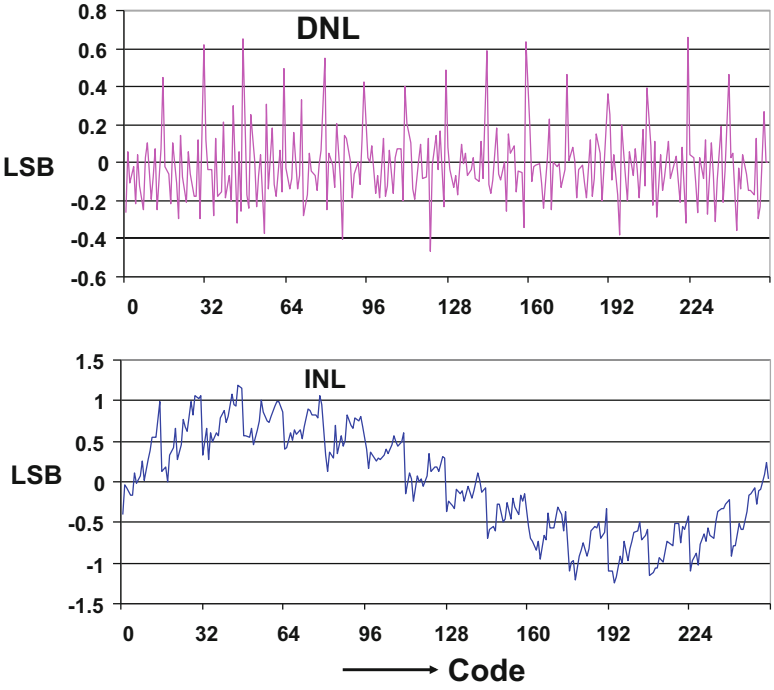


Fig. 4.11 An example of a DNL curve (*upper*) and an INL curve (*lower*) of an eight bit converter

Therefore some systems explicitly demand for monotonic behavior: an increase of the input level must lead to a zero or positive increase of the output code.

In the upper part of Fig. 4.10 an increase of input signal leads to an increase of two LSBs, one digital code is skipped: this error is called "missing code." Note that a missing code is equivalent with a DNL $= -1$ and that is the lowest DNL value possible. Figure 4.11 shows an example of an INL and DNL curve for an eight-bit converter. Both curves give information not only on the performance but also on the architecture of the converter. The saw-tooth shaped pattern in the INL curve indicates a subranging architecture, while the overall "sine" pattern is an indication that a third order distortion is dominant, compare Fig. 4.14. Obviously a rigorous differential design has eliminated the second order component, which would have resulted in a "half sine wave."

Example 4.5. Suppose an INL spec of 1 bit is given, is there a limit to the DNL spec?

Solution. The ideal INL curve is shown in Fig. 4.12 with on each side the curves corresponding to a deviation of ± 1 LSB. The maximum error is artificially created by pulling one trip level 1 LSB to the left and the neighboring trip level one LSB to

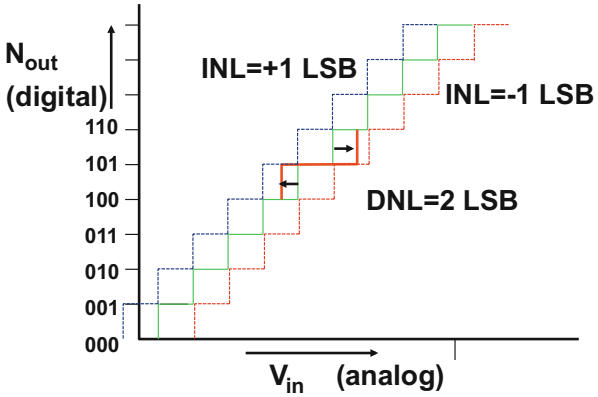


Fig. 4.12 INL curve with INL$= +1$ and INL$= -1$ curves

the right. For clarity the shift is here only 0.9 LSB. The maximum DNL error where the INL stays within ± 1 LSB is therefore +2 LSB.

A more mathematical approach is to split the DNL formula in the difference of two INLs:

$$\mathrm{DNL} = \frac{A(i+1) - A(i)}{A_{LSB}} - 1 = \frac{A(i+1) - A(i)}{A_{LSB}} - \frac{(i+1) \times A_{LSB} - i \times A_{LSB}}{A_{LSB}} =$$

$$\frac{A(i+1) - (i+1) \times A_{LSB}}{A_{LSB}} - \frac{A(i) - i \times A_{LSB}}{A_{LSB}} = INL(i+1) - INL(i) =$$

$$(+1) - (-1) = 2$$

This is one of the classic questions from [2]!

4.5 Modeling INL and DNL

Figure 4.13 shows some typical shapes that can be found in an INL plot of a converter. Mostly a combination of elements is present with some elements more dominating than others. Figure 4.13 (upper part) shows a second and third order distortion curve resulting from transfer curves as shown in Fig. 4.14. These distortions are regular phenomena in all analog circuits. If a sine wave is applied to these converters, the result will inevitably have an INL deviation and contain harmonic distortion.

The starting point for an analysis of the second order shape is an input–output transfer curve of the form $N_{out} \propto V_{in}(1 + \alpha(V_{ref} - V_{in}))$ with $V_{in} = 0, \ldots, V_{ref}$. The maximum deviation of the INL curve from a straight line for an analog-to-digital converter with the INL shape as in Fig. 4.13 (upper, left) is $\alpha V_{ref}^2/4$. Normalizing

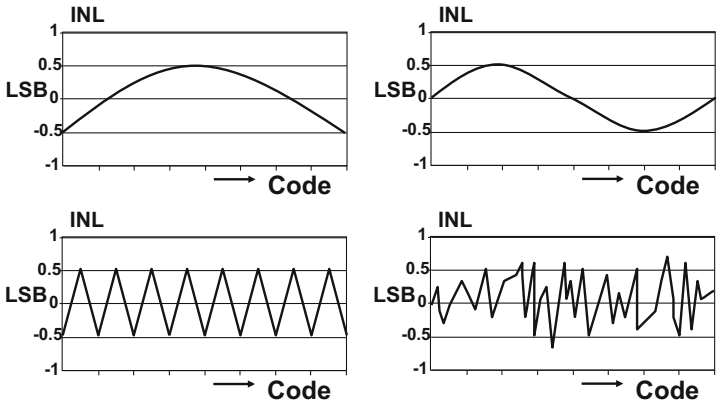


Fig. 4.13 Basic shapes of four typical INL patterns. *Upper curves* show patterns that are caused by second and third order distortion. The *left lower pattern* is often caused to subranging architectures. Next to the depicted *triangle shape*, also saw-tooth patterns can occur originating from binary architectures. The random pattern is typical for flash conversion

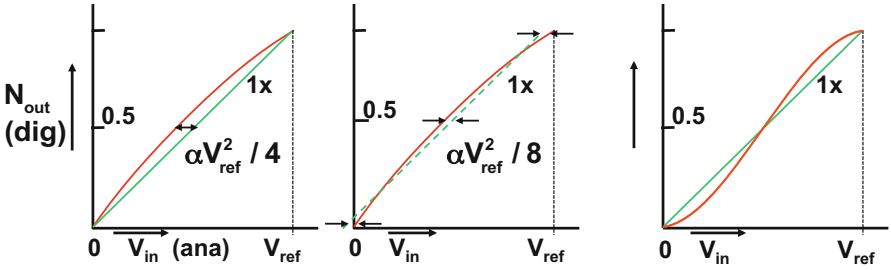


Fig. 4.14 The linear transfer has been mildly distorted by a second order component. *Left*: absolute INL, *middle*: the same error but compared to the best-fitting straight line, the error is split in a positive and negative part. *Right*: third order distortion

this error to an LSB $V_{LSB} = V_{ref}/2^N$ gives

$$\mathrm{INL}_{max} = \frac{\alpha V_{ref} 2^N}{4}$$

in LSBs. In case the INL is presented according to the best-fitting straight line method the curve of Fig. 4.14 (left) is shifted and

$$\mathrm{INL}_+ = \frac{\alpha V_{ref} 2^N}{8} \quad \mathrm{INL}_- = -\frac{\alpha V_{ref} 2^N}{8}$$

The distortion of a full-scale input signal $V_{in}(t) = V_{ref}(0.5 + 0.5\sin(2\pi f_{in} t))$ that is converted via the above curve is calculated using some goniometric equivalences.

The major contributions to the output are found as:

$$\approx \frac{V_{ref}}{2} + \frac{V_{ref}}{2}\sin(2\pi f_{in}t) + \frac{\alpha V_{ref}^2}{8}\cos(2\pi(2f_{in})t)$$

The second order component goes up quadratically if the signal amplitude rises linearly. Consequently the second order distortion component relative to the first order component is

$$\mathrm{HD2} = 20^{10}\log\left(\frac{\alpha V_{ref}}{4}\right)$$

The second order component in an INL plot therefore directly predicts the expected low-frequency distortion. In some textbooks now a simple relation between HD2 and INL is presented:

$$\mathrm{HD2} = 20^{10}\log\left(\frac{\mathrm{INL}_{max}}{2^N}\right) = 20^{10}\log\left(\frac{\mathrm{INL}_{+}}{2^{N-1}}\right)$$

This formulation is valid for the ideal conditions of this derivation. Be careful in using it in arbitrary situations, it just provides an educated guess for the distortion!

The above analysis can be extended to the third order distortion curve as in Fig. 4.14 (right). Approximation of arbitrary sinusoidal and triangular looking shapes in an INL curve can be done with a sine or cosine model for the INL. The periodicity of the argument is twice the number of full periods visible in the utilized input range k. Expanding this function via the Bessel function results in harmonic components in the output at frequencies $(2k-1)f_{in}$, $2kf_{in}$ and $(2k+1)f_{in}$. An input sine wave passing through a converter with a periodical INL leads therefore to higher-order distortion products. In a visual manner these components can be understood as follows: assume the input signal is a perfect triangular shape. When going from 0 to V_{ref} the deviation in the output curve from the ideal replica of the input will be simply the INL curve running from left to right in the time domain. When the triangle input signal returns to complete its period, a mirrored version of the first sequence is added. So one full signal period passes twice through the INL range.

The transfer curve in Fig. 4.13 (lower, right) is typical for a flash converter and array-based digital-to-analog converters. Now the integral and differential non-linearity show the non-uniformity of the trip levels as can be caused by various random processes (e.g., comparator random offsets in full-flash converters). Often it is acceptable to characterize these deviations by a Gaussian distribution. Every trip level is modified by an instance of this distribution:

$$A(i) = i \times A_{LSB} + A_G(i) \tag{4.21}$$

where $A_G(i)$ is a sample from a Gaussian distribution with $\mu = 0$ and variance σ_A^2, modeling the input referred offset of comparator i in, e.g., Volts, see Sect. 8.1.4. The addition of this random process has two consequences: on one hand, the random fluctuation of the trip levels can be seen as a random noise source in series between the input signal and an ideal quantizer. This assumption is valid if the signal varies sufficiently and passes a large number of trip levels. So the noise level at the input of a converter is increased by σ_A^2.

On the other hand, the variation in trip levels is the root cause for the DNL curve and summarized by its maximum value "DNL" in LSBs. The variable $\lambda = \mathrm{DNL} \times V_{LSB}/\sigma_A$ can then be used as the threshold value in the (0,1) Gaussian probability function. Standard tables for a (0,1) Gaussian distribution return the probability that one instance will remain within the interval $(-\lambda, +\lambda)$ corresponding to the probability that the random offset of one comparator remains within the interval (-DNL,+DNL). The yield for $2^N - 1$ comparators is the previously found probability of success raised to the power $2^N - 1$. The maximum resolution that would be used in a flash architecture is 6–8 bit. For this resolution range $|\lambda| = 3.0$–4.0 gives acceptable yields.

If a full-swing sine wave input is assumed, the signal-to-noise ratio for low-resolutions is determined by the quantization errors and DNL errors. Both are mutually independent processes, so their powers $V_{LSB}^2/12$ and σ_A^2 are summed and compared to a full-scale sine wave. For N-bit resolution this yields an update of the $\mathrm{SN}_Q\mathrm{R}$ formula:

$$\begin{aligned} \mathrm{SN}_{Q+DNL}\mathrm{R} &= 10\,^{10}\log 2^{(2N-3)} - 10\,^{10}\log\left(\frac{1}{12} + \frac{\sigma_A^2}{V_{LSB}^2}\right) \\ &= 6.02N - 9.03 - 10\,^{10}\log\left(\frac{1}{12} + \frac{\mathrm{DNL}^2}{2\lambda^2}\right) \end{aligned} \tag{4.22}$$

For $\sigma_A = 0$, DNL $= 0$ the formula results in the well-known $\mathrm{SN}_Q\mathrm{R} = 6.02\mathrm{N} + 1.76\,\mathrm{dB}$.

In Fig. 4.15 the SNR has been plotted in dB versus the DNL for N $= 8$ and a threshold λ of 3.0. The squares indicate the results of Monte-Carlo computer simulations, for both SNR and DNL. The triangles indicate measurement points. At higher DNL the assumption of independence of "DNL-noise" and quantization noise is less valid. Moreover, the simple analysis ignores the fact that sine wave signals do not give uniformly distributed samples, but more often involve quantization errors at the top and bottom than errors around mid range.

Figure 4.15 shows that a DNL $= 0.5$ LSB yields a $-49.2\,\mathrm{dB}$ noise level and a poor DNL $= 1$ LSB results in a SNR $= 47.7\,\mathrm{dB}$, corresponding to a loss of 0.37 ENOB.

Example 4.6. What is the probability that the DNL of the comparators of an 8-bit flash converter will remain within $\pm 3.5\sigma_A$.

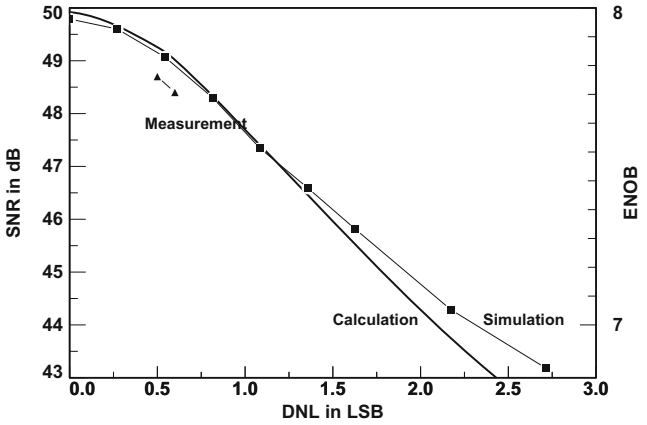


Fig. 4.15 Signal-to-noise ratio for 8-bit quantization with Gaussian-distributed random offset errors versus the expected value of the DNL [258]. The *squares* indicate Monte-Carlo simulations, the *triangles* refer to measurements obtained using the flash A/D converter described in Sect. 8.6.7

Solution. One comparator has a probability to show an input referred offset V_A, with a $(0, \sigma_A)$ Gaussian distribution function. The probability that one comparator remains within the boundaries is found after normalizing to a (0,1) Gaussian distribution:

$$P\left(\frac{-DNL \times V_{LSB}}{\sigma_A} < \frac{V_A}{\sigma_A} < \frac{+DNL \times V_{LSB}}{\sigma_A}\right) = P\left(-3.5 < \frac{V_A}{\sigma_A} < +3.5\right) = 0.999535.$$

For $N = 8$ the probability that all 255 comparators are within the (-DNL,+DNL) interval is $0.999535^{255} = 89\,\%$. With a threshold of $-3.0,+3.0$ the yield drops to 50 %

Example 4.7. A 10-bit ADC is not perfect: at the desired signal and sampling frequency, the DNL is 0.7 bit, while a second order distortion component folds back at a relative amplitude of −56 dB, moreover a fixed clock component at 1/3 of the sampling rate appears at −60 dB. Calculate the effective number of bits of this ADC.

Advise whether the LSB can be removed, so that the succeeding processing runs with 9-bit samples.

Solution. Four components contribute to the loss of signal-to-noise performance: the quantization error, the DNL, the distortion, and the clock component. All these components must be added up in the power domain. First the combined effect of quantization and DNL is estimated with Eq. 4.22 and $\lambda = 3$:

$$\mathrm{SNR_{Q+DNL}} = 6.02 \times 10 - 9.03 - 10\,^{10}\log(\frac{1}{12} + \frac{\mathrm{DNL}^2}{2\lambda^2}) = 60.7\,\mathrm{dB}.$$

Table 4.4 List of powers of unwanted components

$SNR_{Q+DNL} = 60.7\,dB$	$P_{Q+DNL} = 10^{-60.7/10} P_{sig}$	$= 0.85 \times 10^{-6} P_{sig}$
$THD = -56\,dB$	$P_{THD} = 10^{-56/10} P_{sig}$	$= 2.5 \times 10^{-6} P_{sig}$
$Tone = 60\,dB$	$P_{tone} = 10^{-60/10} P_{sig}$	$= 1.0 \times 10^{-6} P_{sig}$
	P_{total}	$= 4.35 \times 10^{-6} P_{sig}$

Table 4.5 List of powers of unwanted components after the 10th bit is removed. The differences with Table 4.4 are high-lighted in bold typesetting

$SNR_{Q+DNL} = \mathbf{55.6}\,dB$	$P_{Q+DNL} = 10^{-55.6/10} P_{sig}$	$= \mathbf{2.75} \times 10^{-6} P_{sig}$
$THD = -56\,dB$	$P_{THD} = 10^{-56/10} P_{sig}$	$= 2.5 \times 10^{-6} P_{sig}$
$Tone = 60\,dB$	$P_{tone} = 10^{-60/10} P_{sig}$	$= 1.0 \times 10^{-6} P_{sig}$
	P_{total}	$= \mathbf{6.25} \times 10^{-6} P_{sig}$

Now all components can be related to the power of the signal, see Table 4.4.

So the SINAD $= {}^{10}\log(4.35 \times 10^{-6}) = 53.6\,dB$ or 8.61 ENOB.

The same effective number of bits can be achieved by a 9 bit converter. What happens after the 10-bit is removed? First, it will be assumed that the errors causing the DNL $= 0.7$ LSB will have the same absolute magnitude. Or on a 9-bit level the DNL $= 0.35$ LSB. Substituting this DNL in the equation gives

$$SNR_{Q+DNL} = 6.02 \times \mathbf{9} - 9.03 - 10\,{}^{10}\log\left(\frac{1}{12} + \frac{\mathbf{0.35}^2}{2\lambda^2}\right) = 55.6\,dB.$$

Now all components can again be related to the power of the signal, see Table 4.5.

Removing the 10th bit causes the SINAD to drop to ${}^{10}\log(6.25 \times 10^{-6}) = 52.0\,dB$ or 8.35 ENOB. Whether this is acceptable, depends on the application. The effect of the tenth bit will be more pronounced at lower input levels: for a 6-dB lower input signal the second order distortion component will be 12 dB or 16× in power reduced, thereby changing the above balance considerably to the advantage of using the tenth bit.

4.6 Figure of Merit

4.6.1 Schreier Figure of Merit

In complex systems and in portable applications the main architectural decisions are often based on bandwidth, resolution, and the available power. Despite the fact that there is no universal law for analog-to-digital power consumption, a practical approach is certainly possible. This approach is based on the observation that more

accuracy or more bandwidth both require more power in a circuit. The combination of bandwidth, power, and accuracy in a Figure of Merit for a circuit allows to compare designs and design choices [59].

The starting point for measuring performances (bandwidth, resolution, and power) is a comparison of the actual power consumed by a circuit $P_{circuit}$ versus the theoretical minimum power. This minimum power level $P_{sig,min}$ is the amount of signal power needed to overcome thermal noise with a certain signal-to-noise ratio (SNR):

$$P_{sig,min} = 4kT \times BW \times \text{SNR} \tag{4.23}$$

$$\text{Power efficiency} = \frac{P_{circuit}}{P_{sig,min}} = \frac{P_{circuit}}{4kT \times BW \times \text{SNR}} \tag{4.24}$$

This relation is often used to evaluate the efficiency of filters, opamps, etc. Some authors use this relation as a starting point for analyzing the lower limits of conversion efficiency, e.g., [60, 61].

This relation is also the basis for the "Schreier Figure of Merit" for analog-to-digital converters[10]:

$$\text{F.o.M.}_S = 10\,^{10}\log \frac{SNR \times BW}{P_{ADC}} = SNR(\text{in dB}) + 10\,^{10}\log \frac{BW}{P_{ADC}}$$

The idea behind this Figure of Merit is that a better SNR requires proportionally lower thermal kT/C noise, leading to proportionally higher capacitor values, that need to be charged with proportionally larger currents. So the power follows the SNR and the F.o.M. remains the same. This allows comparing converters with different specifications and is a basis to judge design quality. At bandwidths (BW) over 100 MHz, the F.o.M. drops, indicating that speed issues, parasitics, etc., become a limiting factor. A simple substitution of the minimum power level leads to the maximum value for the Schreier $\text{F.o.M.}_S = -10\,^{10}\log(4kT) = 198$ dB.

This F.o.M. is designed for comparing thermal noise limited designs, mostly at higher resolution and modest bandwidths. Distortion is ignored, so some variants of this F.o.M. use the SINAD instead of SNR. Figure 4.16 shows the Schreier F.o.M. as a function of the bandwidth. Obviously even the best analog-to-digital converters are still 20 dB or 100x away from the theoretical limit. For lower bandwidths the best converters reach an F.o.M. of 175–178 dB.

[10]The author of this book was educated with the notion that "log" operations can only be performed on dimensionless quantities. Obviously this is not the case here.

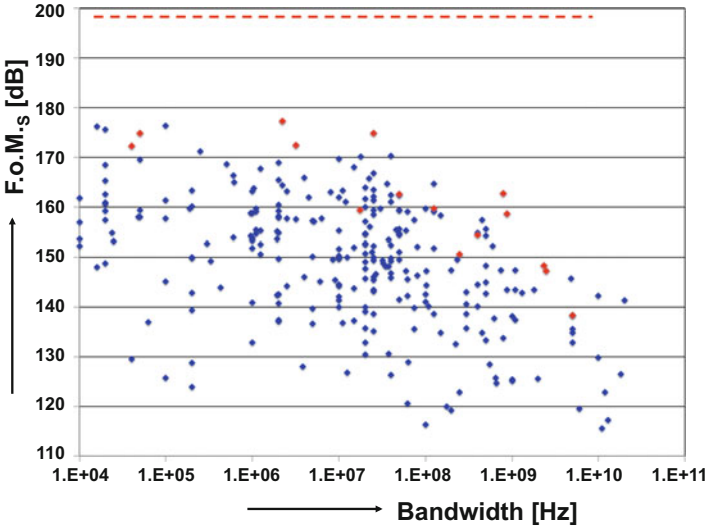


Fig. 4.16 The Schreier Figure of Merit for various analog-to-digital converters as published on the International Solid-State Circuits Conference. The 2015 edition has been marked with *red symbols*. (From: B. Murmann, "ADC Performance Survey 1997–2015," [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html)

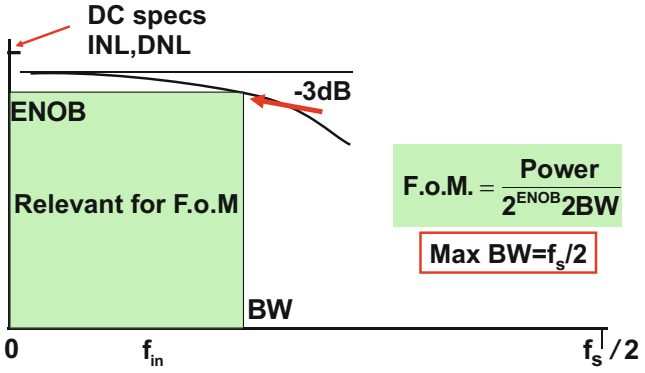


Fig. 4.17 Definition of the combination of bandwidth and resolution for the Figure of Merit

4.6.2 Walden Figure of Merit

Figure 4.17 shows the behavior of the effective number of bits (comprising signal-to-noise-and-distortion) as a function of the applied input frequency. At higher frequencies the conversion becomes less accurate due to the increased distortion and other unwanted effects.

The Walden F.o.M. [59] measures the performance of a converter as a combination of bandwidth BW and resolution ENOB, chosen such that ENOB is roughly 3 dB or 0.5 effective bits below its low-frequency value. As the shape of this curve in this signal-frequency range often is dictated by a first order function, the product of resolution ($\propto 1/\omega$) and bandwidth ($\propto \omega$) is not sensitive to the actual choice. The maximum bandwidth is limited to half of the sample rate due to Nyquist theorem. This combination of power, bandwidth, and the associated measured resolution are the ingredients for the Walden Figure of Merit:

$$\text{F.o.M.}_W = \frac{\text{Power}}{2^{ENOB} \times \text{Minimum}(2BW, f_s)} \quad \text{[Energy per level]} \tag{4.25}$$

In this F.o.M. a factor of $2BW$ is used to allow comparison with older F.o.M. numbers that use the nominal resolution N and the sample rate f_s. A low F.o.M. indicates that a converter uses less power for a certain measured specification, or delivers a better specification for the same power.

The differences between converters are attributable to architecture (number of comparators), technology, noise or limited matching, poor design, etc.

The term 2^{ENOB} replaces the signal-to-noise power ratio in Eq. 4.24. Essentially $2^{ENOB} \propto \sqrt{SNR}$ but gives a better fit of the available data on analog-to-digital converters, see Figs. 4.18 and 4.19. One reason is that ENOB contains all unwanted

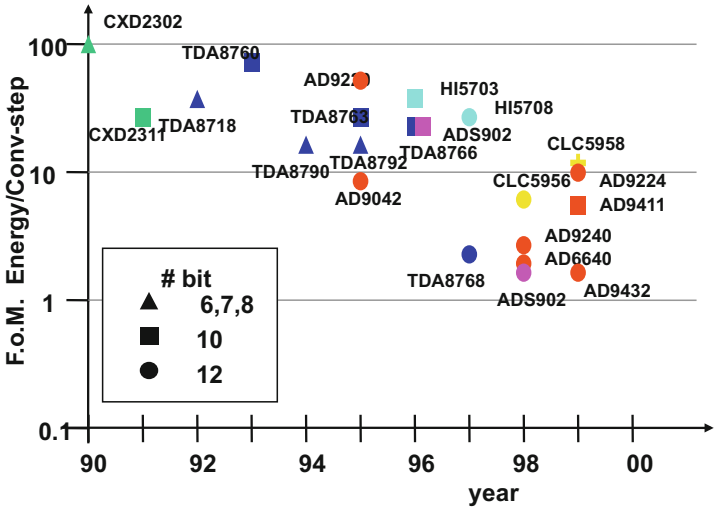


Fig. 4.18 Figure of Merit for various industrial analog-to-digital converters in the time frame 1990–2000

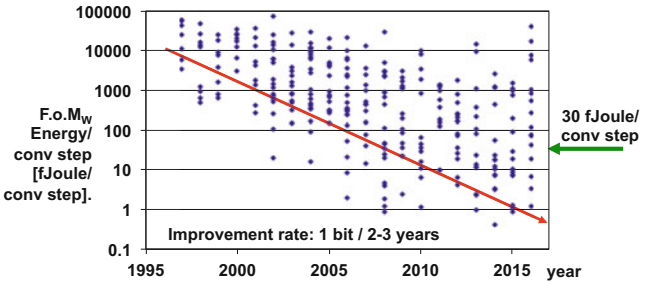


Fig. 4.19 Figure of Merit for various analog-to-digital converters as published on the International Solid-State Circuits Conference. (From: B. Murmann, "ADC Performance Survey 1997–2016," [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html)

signal aberrations, not only thermal noise. Another part of the explanation can be found in the observation that the theoretical thermal noise formula allows any choice of currents and voltages needed to form the power. In practice the voltage choice is restricted to the available power supply, which may explain a better data fit when using ENOB for designs that are not thermal noise limited.

These F.o.Ms. allow to compare various converter principles and resolution/bandwidth combinations. If the F.o.M is plotted over time as in Figs. 4.18 and 4.19, a clear reduction of power for a certain specification is observed. This plot implies a rate of 1 bit per 3 years improvement.[11]

Next to a comparative function the F.o.M. can also be used to predict the conversion power for a specific architecture choice. In the year 2016 an efficient converter uses according to Fig. 4.19 less than 30 fJ per conversion step. This result is obtained by comparing various analog-to-digital converter architectures in various stages of industrialization. It may be useful to limit to just one architecture and compare equivalent stages of development.

This value of F.o.M.= 30 fJ/conv can now be used to calculate the allowable power for a design target. Of course this estimate is based on some crude assumptions and is merely an indication for the order of magnitude that one can expect.

$$\text{Estimated power} = \text{F.o.M.} \times 2BW \times 2^{\text{ENOB}} \tag{4.26}$$

[11] Compared to Moore's law for digital circuit where speed doubles and area and power halves for every generation (2 years) this is a meager result.

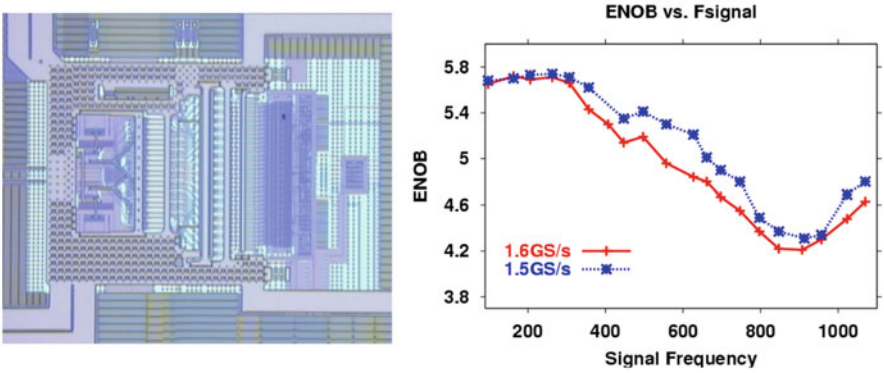


Fig. 4.20 Photograph and performance plot of a 6-bit 328-mW 1.6-Gs/s flash ADC

Figure 4.21 shows the projected power dissipation in a field spanned by a resolution and bandwidth axis. Moving the analog-to-digital conversion from, e.g., direct telephony speech level to the GSM baseband digital level not only means a shift from (8b/3 kHz) to (12b/200 kHz), but also costs three orders of magnitude in power consumption.

Example 4.8. Determine the Walden F.o.M. for the ADC reported in [199], Fig. 4.20.

Solution. The reported power is 0.328 W. A suitable data point on the plot of the ENOB versus the input frequency is 5.4 bit at 400 MHz. Yielding

$$\text{F.o.M.} = \frac{\text{Power}}{2^{ENOB} \times \text{Minimum}(2BW, f_s)}$$

$$\text{Energy per level} = \frac{0.328}{2^{5.4} \times 2 \times 400 \times 10^6} = 9.7 \times 10^{-12}$$

Some more examples are given in Table 4.6.

Example 4.9. Give a power estimate for a 100 MHz bandwidth, 14 ENOB analog-to-digital converter.

Solution. With a F.o.M.= 30 fJ/conv step, the estimated power is F.o.M. $\times 2^{14} \times 2 \times 10^8 = 0.1$ W (Fig. 4.21).

Table 4.6 Power efficiency of some ISSCC published analog-to-digital converters

Year	Author	Architecture	N bit	f_s Ms/s	Power W	BW MHz	ENOB bit	F.o.M$_W$. fJ/conv
2006	Schvan [195]	TI Flash	6	24000	1.2	12000	3.5	4400
2013	Verma [299]	TI Flash	6	10300	0.24	5000	5.1	149
2006	Geelen [231]	Pipe	10	100	0.035	50	9.3	550
2009	Brooks [242]	Pipe	12	50	0.0045	25	10.0	90
2012	Chai [224]	Pipe	10	200	0.0054	100	9.5	46
2007	Hsu [288]	TI pipe	11	800	0.35	400	8.7	1100
2013	Setterberg [301]	TI pipe	14	2500	23.9	1250	9.8	10427
2006	Shimizu [211]	Subrange	12	40	0.03	20	10.5	520
2007	Hesener [256]	SAR	14	40	0.066	0.96	13.8	2200
2007	Craninckx [246]	SAR	9	50	0.00029	10	7.4	65
2008	Elzakker [248]	SAR	10	1	0.0000019	0.5	8.5	4
2014	Harpe [253]	SAR	12/14	0.032	$3.5\ 10^{-7}$	0.016	11.3	4.5
2011	Doris [283]	TI SAR	10	2600	0.48	1300	7.6	849
2015	Lim [165]	Pipe/SAR	13	50	0.001	25	11.5	4
2015	Brandilini [298]	TI Pipe/SAR	10	5000	0.15	2500	7.4	45
2009	Naraghi [264]	Linear	9	1	0.000014	0.5	7.9	100
2013	Chae [327]	Incremental	20	0.05	$6\ 10^{-6}$	12.5^{-6}	19.6	314
2007	Christen [323]	$\Sigma\Delta$	12	240	0.021	10	10.2	900
2006	Schreier [355]	BP $\Sigma\Delta$	15	264	0.375	8	12.3	4500
2011	Bolatkale [339]	TC $\Sigma\Delta$	11	4000	0.256	125	10.5	705
2012	Shettigar [343]	TC $\Sigma\Delta$	14	3600	0.015	3.6	11.5	73
2014	Dong [336]	TC $\Sigma\Delta$	12	3200	0.235	50	11.8	3.4

Some care must be taken in comparing the numbers as the measurements conditions are not identical

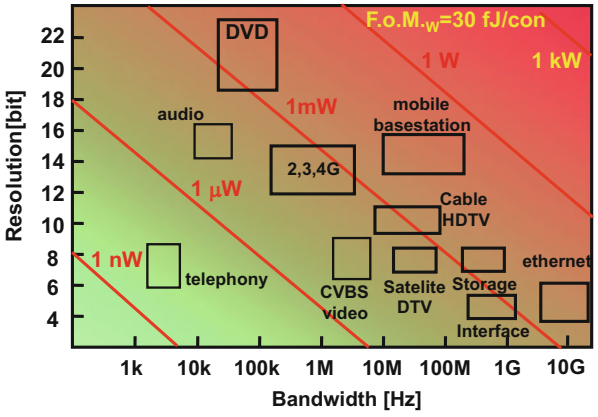


Fig. 4.21 In this bandwidth/resolution field the global specifications of some consumer and communication analog-to-digital interfaces are shown. Also a power estimation based on an F.o.M.$_W$ of 30 fJ/conv step is indicated

4.6.3 Digital-to-Analog Converter Figure of Merit

Several F.o.Ms exist to compare high-performance current-steering digital-to-analog converters, see Sect. 7.3.2. In [62] a relation between peak–peak signal swing V_{pp}, signal output frequency f_{out}, spurious free dynamic range ratio (SFDRR), and total consumed power P_{DAC} is proposed as F.o.M.:

$$\text{F.o.M.} = \frac{V_{pp} f_{out} \text{SFDRR}}{P_{DAC}} \tag{4.27}$$

where the SFDRR equals $10^{\text{SFDR}/20}$. A comparison with the F.o.M. for analog-to-digital converters in Eq. 4.25 shows that an additional term V_{pp} accounts for the power (and design problems) needed for generating the analog signal in a digital-to-analog converter.

4.6.4 Nyquist and Oversampled Architectures

From the analysis leading to Eq. 4.7, two main options emerge to reduce the quantization error power in a given bandwidth: increase the resolution N (so A_{LSB} becomes smaller), or spread out the noise thinner over more bandwidth by increasing f_s. Increasing the resolution by one bit gives a 6-dB higher SN_QR. Quadrupling the sampling rate divides the quantization power per Hertz by four and results in a fixed bandwidth in a 6-dB higher SN_QR, or 1 ENOB.

These two directions correspond to the main classification of analog-to-digital converter architectures:

- In Nyquist converters the signals use most of the available bandwidth between 0 and $f_s/2$. Their resolution is increased by generating a denser pattern of quantization levels by means of capacitor banks, resistor strings, etc. Technology limitations such as component mismatch limit the achievable resolution to $N =$ 12–14 bit. Calibration can help, but at the cost of area, time, and power. Generally the bandwidth of interest in Nyquist converters is high: close to $f_s/2$ depending on anti-alias filter needs. Their efficiency measured as $P/2f_{s,Nyquist}$ is good as the bandwidth related power dominates over the power in various biasing, reference, and support circuitry. Interfacing these converters to the digital world is trivial except where extreme sample rates are used. The issues associated with handling extremely high-speed digital data stream are beyond the scope of this book. Chapters 7, 8, and 9 discuss Nyquist digital-to-analog and analog-to-digital converters.
- The quantization error spectral density can be lowered for a given resolution N by increasing the sample rate f_s. Oversampled converters such as noise shapers and sigma-delta converters use the available large difference between the bandwidth of interest and the technologically available sampling speed to spread out the

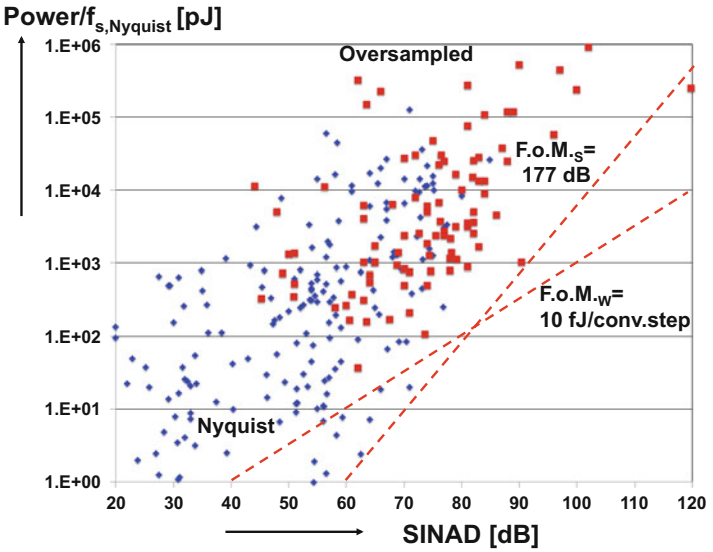


Fig. 4.22 The power efficiency (power per bandwidth) versus the SINAD. Obviously oversampled converters populate the high SINAD range, while Nyquist converters are more efficient. (From: B. Murmann, "ADC Performance Survey 1997–2015," [Online] Available: http://web.stanford.edu/~murmann/adcsurvey.html)

quantization error power. Now the resolution comes from the time domain. Some form of glue circuitry (e.g., decimation) is required to connect these converters to the digital processor. Chapter 10 discusses these converters.

Figure 4.22 compares both architecture directions.

Exercises

4.1. Suppose a DNL spec of 1 bit is given, is there a minimum or maximum limit to the INL spec?

4.2. The output signal of an FM intermediate frequency circuit has a bandwidth of 100 kHz at a frequency of 10.7 MHz. An analog-to-digital converter samples this signal at 5.35 Ms/s. What resolution is needed to obtain an SNR due to quantization of 14 bit in 100 kHz bandwidth.

4.3. A white noise dither signal is applied with an in-band root-mean-square level of $0.289 \times V_{LSB}$. Give an expression for the best obtainable SNR.

4.4. A 14-bit ADC has to convert a 2 $V_{peak-peak}$ signal at a quality level corresponding to ENOB=12.5 bit. The sampling capacitor is 1 pF. How much in-band distortion can you tolerate?

4.5. A signal-to-noise ratio of 90 dB is required in a 20 kHz bandwidth. In the system a 650 kHz sampling rate is available and a 650 MHz rate. Which combination of resolution and sample rate can be used to achieve the desired SNR? Which converter has your preference?

4.6. What is the maximum spurious free dynamic range of a perfect 8-bit analog-to-digital converter.

4.7. What is the maximum allowable INL to reach a spurious free dynamic range of 80 dB with a 12-bit converter.

4.8. A sine wave is quantized with 2-bit. Is it possible to find an amplitude of the sine wave that will result in zero third order distortion?

4.9. A sine wave of 10 kHz is sampled at 60 ks/s and quantized with 8-bit resolution. Make a drawing of the resulting waveform. Is this a favorable situation? What happens if the sine wave contains odd harmonics. Draw the spectrum.

4.10. An INL curve of a 10-bit digital-to-analog converter shows a value of -0.4 LSB at the ends of the range and at $+0.4$ LSB in the middle. Give an estimate for the second order distortion ratio.

4.11. Derive a relation between the INL and the third order distortion in Fig. 4.13.

4.12. A full-range sine wave is quantized by an analog-to-digital converter with an INL curve as in Fig. 4.13 (lower, left). At what frequencies do you expect harmonic components. Now the sine wave is 3/4 range, at what frequencies will now harmonic components appear?

4.13. A 6-bit converter has a transfer characteristic described by $y = x + x(x - 0.25)(x-0.75)(x-1)$ where $x = 0 \ldots 1$. Determine the INL curve and the harmonics if this converter is driven by a signal $x(t) = 0.5 + 0.5\sin(\omega t)$. Repeat with a signal of $x(t) = 0.5 + 0.25\sin(\omega t)$.

Chapter 5
Accuracy

In any analog-to-digital converter the time- and amplitude-continuous signal is mapped and rounded to a grid defined by discrete time moments and discrete amplitude levels. Crystal clocks or high-quality phase-locked loops generate the time moments. A discrete amplitude scale is obtained by dividing a reference quantity in LSB portions: A_{LSB}. A proper design will use similar components to get accurate amplitude levels. If all components are made of the same material, have the same size, orientation, etc., the absolute value of these components will not matter, only their ratio. A resistor string of 1024 resistors connected between two voltages will perfectly subdivide the voltage difference, irrespective of whether a single resistor measures $10\,\Omega$ or $10\,\mathrm{k}\Omega$. Of course power and speed parameters are still dependent on the absolute values.

When approaching atomistic dimensions, however, the idea of perfect ratios between equally designed groups of components starts to vanish. Equally designed components show random behavior or "variability" that affects the perfect ratios and consequently disturbs the conversion.

In this chapter various deterministic and random effects on the accuracy of the converter are discussed.

5.1 Variability

Variability[1] is generally interpreted as a collection of phenomena characterized by uncontrolled parameter variation between individual transistors or components in a circuit. This collection is populated with a large number of effects ranging from offset mechanisms to reliability aspects.

[1]The following sections are an adaptation of Chap. 15 "Modeling of MOS matching" in: "Compact Modeling: Principles, Techniques and Applications" by Gildenblat (ed.) [12]. The original publication is available at www.springerlink.com and was co-authored with Hans Tuinhout and Maarten Vertregt.

M. Pelgrom, *Analog-to-Digital Conversion*, DOI 10.1007/978-3-319-44971-5_5

Table 5.1 Deterministic and statistical effects

Deterministic	Pseudo-statistical	Statistical
Process	Substrate noise	Line edge roughness
Supply voltage	Dyn IR drop, jitter	Noise (1/f, kT)
Temperature	Temp. gradient	Granularity
Offsets, IR drop	Cross-talk	Dopant fluctuation
WPE, Stress, STI	Substrate noise	Mobility fluctuation
Proximity effects	CMP density	Interface states
Soft breakdown	Drift	Work function fluctuation

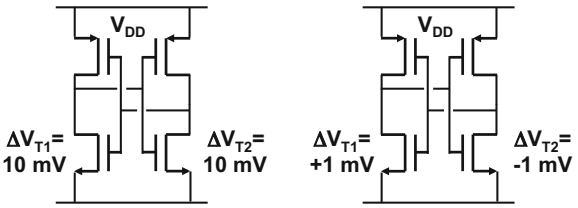


Fig. 5.1 Two latches in a meta-stable condition experience global and local offsets

From a statistics point of view these (time-independent) effects can be subdivided in two classes: deterministic and stochastic and in designer's terms: offsets and random matching. As simple as this division seems, there is a complication: a number of phenomena are from a physics point of view deterministic, but due to circuit complexity, a statistical approach is used to serve as a (temporary) fix. An example is wiring stress, where the complexity of concise modeling is too cumbersome, see Table 5.1.

From a philosophical perspective, the listed random effects are not truly stochastic effects. Here a practical point of view will be used: all effects that are reproducible from die-to-die will be categorized as deterministic. If the variation source changes the behavior of every individual device with respect to the average, the effect will be described with stochastic means.

Example 5.1. The latch in Fig. 5.1 (left) is in a meta-stable position. Due to a global change in back-bias voltage the threshold voltage of the NMOS transistors shifts by 10 mV. Another meta-stable latch(right) experiences cross-talk from neighboring wiring equivalent with a positive, respectively, negative shift on both nodes. What will happen with these two latches?

Solution. Global excitations and variations are canceled by differential structures such as the differential latch topology of Fig. 5.1. The global change of threshold voltage will not impact the latch and the meta-stable condition will continue.

However, local variations like asymmetrical cross-talk, will also affect differential circuits. Now the latch will amplify the cross-talk signal and flip into a 1-0 state.

5.2 Deterministic Offsets

Deterministic or systematic offsets between pairs of resistors, CMOS transistors, or capacitors are due to electrical biasing differences, mechanical stress, lithographic, and technological effects in the fabrication process [63].

It may seem trivial, but good matching requires in the first place that matched structures are built from the same material and are of equal size. This implies that a 4.5:1 ratio is constructed as a 9:2 ratio, with 11 identical elements, arranged in a common centroid, see Figs. 5.26 and 5.2. The required 4.5:1 ratio holds for every aspect of the combination: area, perimeter, coupling, etc. Any global variation will scale and the same ratio is maintained. This technique is applied in capacitive filters or binary weighted capacitor banks, digital-to-analog conversion elements, etc. Also larger ratios or less trivial integer ratios like 22/7 for π are preferred for matching over simply multiplying the width by 3.14. In case the required ratio cannot be approximated sufficiently within the boundary conditions, an optimum ratio is chosen and one cell is increased or reduced in size to obtain the required ratio. Ratios that approximate an integer (3.95, 4.15) are constructed by modifying one element.

A pitfall can occur when an existing lay-out is mathematically scaled (e.g., all dimensions multiplied by 0.87) and the resulting dimensions rounded off to fit the new lay-out grid. The rounding operation can result in unexpected size and ratio deviations in originally perfectly matched devices.

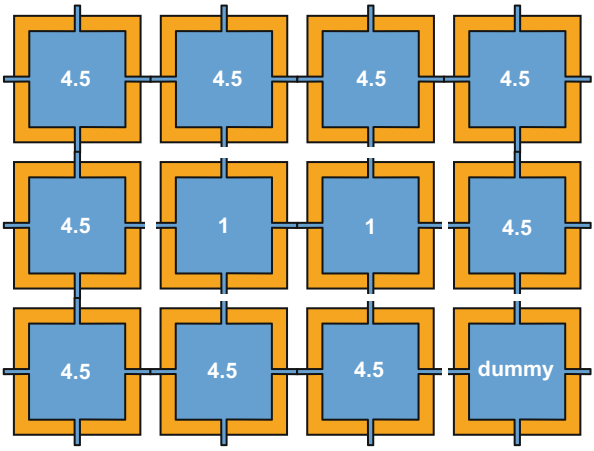


Fig. 5.2 Example of a 1:4.5 capacitor ratio, laid-out with 11 units. For symmetry reasons one dummy capacitor is left in the lower right corner. Most lay-outs will surround this structure with a ring of dummies resulting in a 5×6 matrix of capacitors

5.2.1 *Offset Caused by Electrical Differences*

For electrical matching, the voltages on all the elements must be identical. Node voltages are affected by voltage drops in power lines, leakage currents in diodes, substrate coupling, parasitic components, etc. On a building block level DC-offset can be introduced by phenomena as self-mixing, see Sect. 2.3.1. Deterministic effects can indirectly be caused by electrical currents or voltages, e.g., heat gradients due to power dissipation and temporary charging effects due to phenomena as negative bias temperature instability [64].

During the design of capacitors, electrical fields of other conductors can play a role. In Sect. 3.3.1 and in Fig. 7.46 some remarks are made specifically for the design and lay-out of capacitors.

Figure 5.3 shows a surprisingly often encountered example of electrical mismatch: a pair of transistors with a wire running along the sources. The current I_{GND} in this wire will cause a voltage drop between the two source terminals of the transistors. This voltage drop shows up as an offset ΔV_{GS} in the gate-drive voltage. Such problems increase when additional currents are routed via this wire. A star-like connection is well-suited to avoid this problem. A rigorous inspection of the lay-out in which the main current paths have been identified is always needed when offset problems are suspected.

Example 5.2. The NMOS transistors in Fig. 5.3 have dimensions of 20/0.2 in a 0.18 µm process and are biased with 0.2 V drive voltage. The sources are connected with a 0.2 µm wide metal wire of 0.1 Ω/□. How much offset will occur?

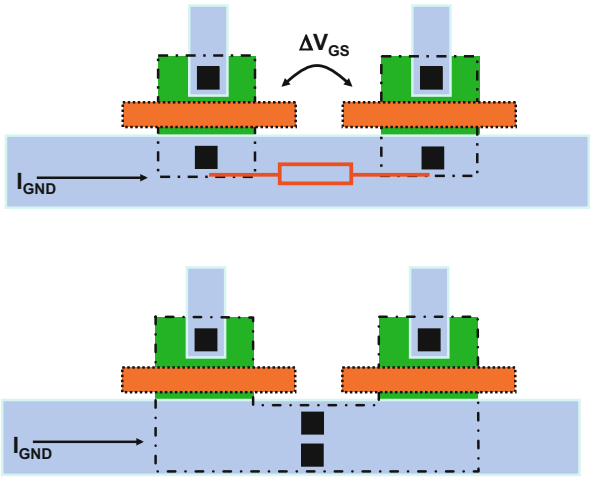


Fig. 5.3 A small resistance in the source connection wire will result in offset in the gate voltage. A common connection point is better from an offset point of view

Solution. With a current factor of $\beta_{\square} = 300\,\mu A/V^2$, see Table 4, each transistor will conduct a current of 0.6 mA. Between the sources there is some 25 µm distance, resulting in a resistance of 12.5 Ω. So the voltage offset is 7.5 mV translating into 18 µA of current difference.

5.2.2 Capacitors

In CMOS technology a few options are present for the implementation of the capacitor:

- Diffusion capacitances are not suited for most specifications. Leakage, non-linearities, and low capacitance per unit area render this device unattractive.
- The gate-to-channel capacitor gives the highest capacitance per unit area. However, this device requires a significant turn-on voltage in excess of the threshold voltage. Using voltages below the threshold voltage is not possible as the inversion layer will disappear. Good for decoupling bias and power lines, but then a reliability hazard.
- In older processes a double polysilicon capacitor option is offered, see Fig. 5.4. Top contacts are normally not allowed as the contact hole construction can damage the plate-to-plate insulator. A large signal can lead to the depletion of the polysilicon resulting in voltage non-linearities. Also there is a resistive element when large amounts of polysilicon structures are connected.
- In some process variants the so- called metal-insulator-metal capacitance option is present. The special dielectric has excellent linearity properties.
- Interconnects allow to design plate capacitors and fringe capacitors. Stacking various layers of interconnect, where odd and even numbered stripes and layers form the plates of a capacitor, is generally a good solution. This capacitor requires no bias, has a good linearity and low parasitics.

Designers will typically choose for a fringe or metal-plate capacitance in an advance process or a double-poly in an older process. The horizontal plate-capacitor

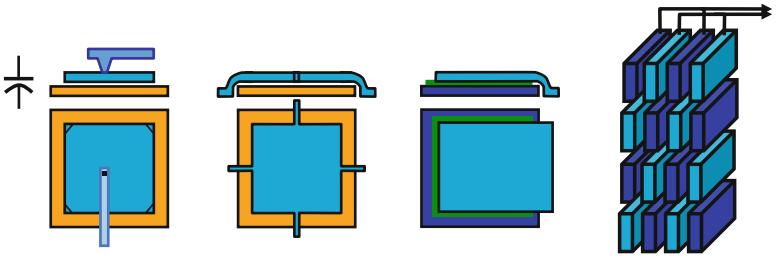


Fig. 5.4 Some common implementations of capacitors. *From left*: double poly with top contact, double poly with side contacts, metal-insulator-metal, and fringe capacitors

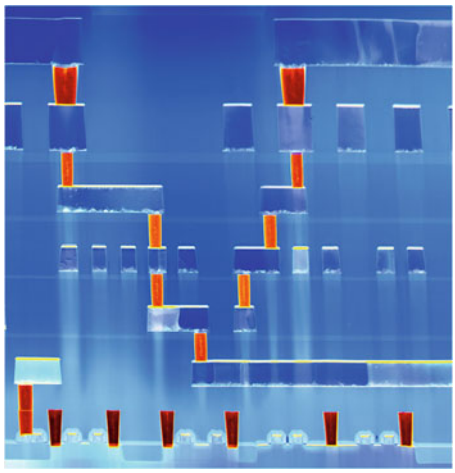


Fig. 5.5 A cross-section through a 65/90-nm CMOS die. Clearly the wiring dominates over the transistors at the bottom (Courtesy: Philips/STM Crolles alliance)

constructions show some form of asymmetry because one layer is closest to the substrate. This type of capacitor has a large bottom-plate parasitical capacitor and nearly no top-plate parasitic. In the capacitor symbol a curved bottom plate indicates the terminal with the large parasitical component. Depending on the circuit topology this capacitor terminal also creates susceptibility to, e.g., substrate noise. Next to that also horizontal coupling must be considered. Clocked interconnect lines should be kept away of the capacitor or even shielded by placing grounded lines in between.

Figure 5.5 shows the cross-section through a deep sub-micron technology die. The transistors are just visible at the bottom of the photo. The wiring dominates. The height of the wires and the isolation thickness make the horizontal coupling of narrow wires more important than the traditional vertical plate-to-plate contribution. Simple CAD analysis (and process characterization) supports only this vertical component. From partly overlapping structures only the vertical component is analyzed. More advanced tools include horizontal coupling, and coupling to structures in the vicinity. Full electrostatic analysis requires special tools and can only be executed on small structures in a reasonable amount of CPU time.

A simple view on a capacitive structure is shown in Fig. 5.6: the top plate has more couplings than just to the lower plate. Instead of relying on CAD extraction tools, it is better to design the capacitors and sensitive signal wires in a manner that defines by lay-out what coupling can exist.[2] Figure 5.6 (right) shows an often used technique where the sensitive plate is completely surrounded ("caged") by the bottom plate. Also transmission lines can be designed in this way.

In successive approximation converters, Sect. 8.6, sometimes very small capacitors are needed. In this case, inside one metallization layer a finger is surrounded by a u-shape, see Fig. 5.7.

Table 5.2 summarizes various capacitor characteristics.

[2]It is nearly always better to rely on simple and oversee-able structures than on the ability to extract precisely the parasitics.

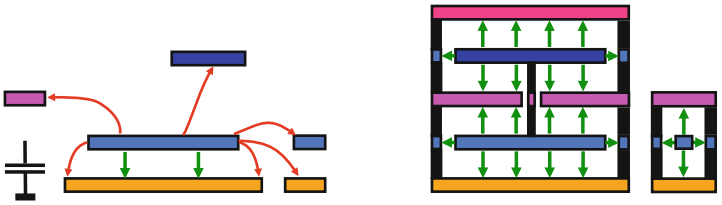


Fig. 5.6 Building a capacitor in an advanced technology requires to inspect all possible couplings between wires. In this cross-section the upper plate does not only couple to the plate below, but field lines escape in all directions and find other conductors to land on. *Right*: the field lines in "caged capacitors" and caged wires can only land on directly surrounding metal, e.g., [164]

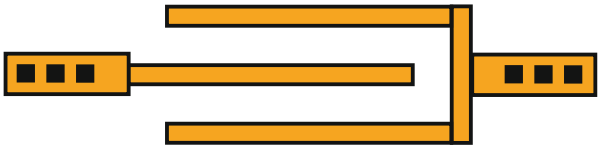


Fig. 5.7 Very small capacitors (<1 fF) are often realized inside one layer, this is a lay-out top view. For example, [82]

Table 5.2 An indication of capacitor characteristics in a semiconductor process ranging from 0.18 μm to 90 nm generations (sources, e.g., ITRS [13] and various publications)

Material stack	Capacitance fF/μm^2	Voltage coeff V^{-1}	Temp coeff °C^{-1}	Matching $\%/\sqrt{\text{fF}}$
Diffusion	0.5		3×10^{-4}	
MOS gate 0.18 μm	8.3	$3–5 \times 10^{-2}$		
Fringe capacitors	1.5			0.3
MIM capacitors	4–15	10^{-5}		0.3
Plate capacitors($d_{ox} = 80$ nm)	0.43/layer			0.5
Poly-poly 0.35 μm	0.8	$5–10 \times 10^{-4}$	-8×10^{-5}	5

5.2.3 *Resistors*

In sub-micron CMOS processes the designer has a choice between polysilicon and diffused layers to design a resistor or string of resistors. Metallization layers and well diffusions are not recommended: too much variation. A polysilicon layer has less parasitic capacitance and a low voltage coefficient, see Table 5.3. A diffused layer gives a better matching performance. The voltage dependency of the diffused resistor in Fig. 5.8 is canceled by placing the resistor in a well of opposite doping and biasing this well with the same voltage difference.

Figure 5.9 shows two lay-out examples of strings of resistors. Note the dotted line that indicates the mask used for removing the low-ohmic salicide layer on top of the resistors. The typical resistance of diffused and polysilicon material as specified by this mask, is of the order of 100 Ω per square. The salicide reduces

Table 5.3 An indication of resistor characteristics in a semiconductor process (0.18 µm to 90 nm generations) from ITRS [13] and various publications

Material	Square resistance $\Omega/\square$	Voltage coeff V^{-1}	Temp coeff $°K^{-1}$	Matching A_R %µm
n/p diffusion	75..125	1×10^{-3}	$1..2 \times 10^{-3}$	0.5
n well diffusion	1000	80×10^{-3}	4×10^{-3}	
n-polysilicon	50..150		$-1.. + 1 \times 10^{-3}$	2
p-polysilicon	50..150		0.8×10^{-3}	2
Polysilicon(silicide)	3..5		3×10^{-3}	
Aluminum	0.03..0.1		3×10^{-3}	

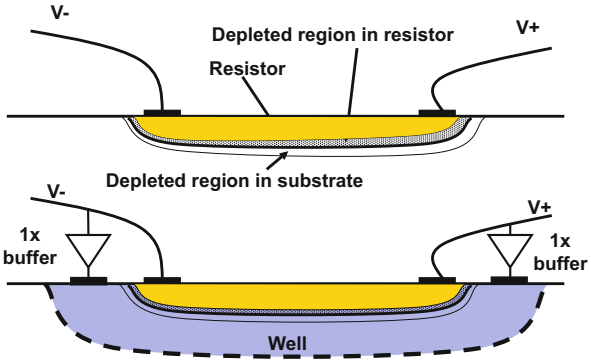


Fig. 5.8 Due to the voltage difference over a diffused resistor, the active resistor is thinner at higher voltages. Placing the resistor in a well of opposite dope and biasing the well with the same voltage difference avoids the voltage dependence

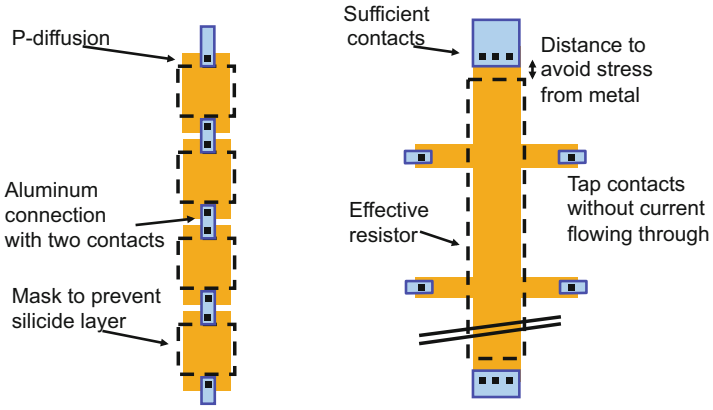


Fig. 5.9 On the *left-hand side* is a ladder build up with discrete resistors. On the *right-hand side* a ladder structure is shown for high accuracy, see Sect. 5.2

this resistance to a few Ohms. The left-hand construction uses a fixed resistor lay-out connected by wiring. This is not an optimum construction. The current has to pass through the contacts. In many processes contact areas show a lot of additional variation in resistance and stress due to the metals and oxide borders. On the right-hand side a construction is shown where the tap voltage connections do not carry current. Any variation in position, resistivity of the contact, etc., is not relevant. The current supplying connections to the ladder are designed with sufficient contacts and are placed at some distance to reduce the material stress caused by the presence of aluminum wires, see also Fig. 5.25. Preferably a few dummy taps are inserted between the main connections and the first relevant taps.

This construction cancels a number of effects, such as contact hole resistance. For large ladder structures also attention must be paid to gradients. Due to processing or heat sources the resistivity of the ladder material is not constant with distance. Gradients will create a non-linear voltage distribution over the ladder, Fig. 5.10 (left). If a ladder consists of 2^N resistors each nominally of value R with a gradient defined by a difference of ΔR between two adjacent resistors, then the value of the i-th resistor ($i = 1, .., 2^N$) can be described as $R(i) = R + (i - 2^{N-1})\Delta R$. This model places the nominal resistor in the middle at position $i = 2^{N-1}$. The current flowing through the ladder is constant, so the voltage on a node $m = 0, .., 2^N$ can be described by a resistor ratio:

$$\frac{V(m)}{V_{ref}} = \frac{\sum_{i=0}^{i=m} R + (i - 2^{N-1})\Delta R}{\sum_{i=0}^{i=2^N} R + (i - 2^{N-1})\Delta R} \approx m2^{-N}(1 - 2^{N-1}\frac{\Delta R}{R}) + m^2 2^{-N-1}\frac{\Delta R}{R} \quad (5.1)$$

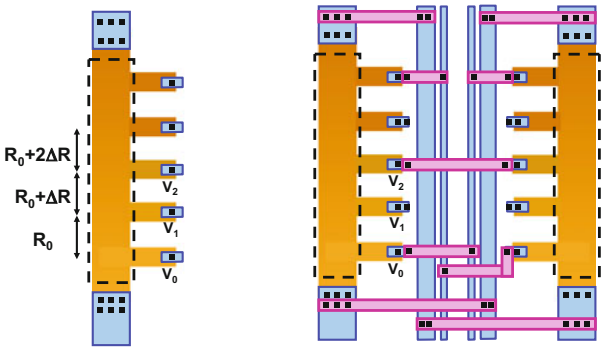


Fig. 5.10 *Left*: a resistor ladder with gradient. *Right*: A cross-coupled ladder structure. The drawn contacts serve merely as an indication, in a real lay-out large arrays of contacts will be used

and $V(m = 0) = 0$. This formula is of the form: $y = x + ax^2$. Therefore the distortion and INL calculations of Sect. 4.5 apply.

Cross-coupling as shown in Fig. 5.10 (right) eliminates gradients. In this structure a second ladder is placed in the lay-out parallel to the original ladder. However, from an electrical perspective this ladder is connected upside-down to the first ladder. Only the two extreme connections carry (large) currents. The currents in the intermediate connections are ideally zero, but will never be large.

Tuinhout et al. [84] is one of the few dedicated papers on resistor quality. The observation here is that a quality level of 12–14 bit is reachable.

5.2.4 *Offset Caused by Lithography*

During the lithography process the structures drawn in the lay-out are transferred to a mask and in the resist. Accuracies on masks in advanced processes range from 1 to 5 nm on silicon.[3] In the next step physical structures are etched into a wafer. During this process there are many crucial details that will affect the quality of the patterning. Figure 5.11 shows the frequencies of free-running ring oscillators of

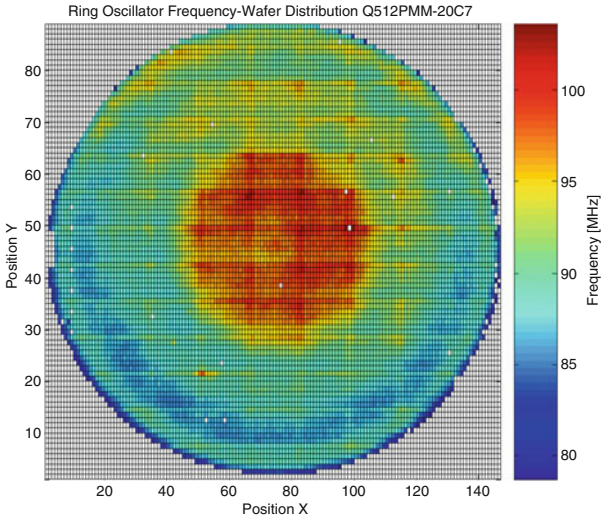


Fig. 5.11 A wafer from a 90-nm CMOS production lot on which a free-running oscillator frequency is measured. The frequency deviation (80–100 MHz) is largely an indication for the variation in gate-length due to mask-plate dimensional errors, lithography, and the processing steps that determine the electrical gate-length (Courtesy: B. Ljevar, NXP)

[3] Price is the determining factor.

circuits that are measured on a 300-mm wafer. Ring oscillators are most sensitive to gate-length variation, so the frequency map indirectly represents a gate-length map. Most prominent is the middle area, where a 20 % higher frequency is measured with respect to the rest. A potential cause is the heat non-uniformity during the high-temperature dopant activation step or the specific gas flow in the etch equipment. A second phenomena is visible as a pattern of rectangular shapes. The wafer stepper uses a reticle containing 16 by 7 devices. This $\approx 400\,\mathrm{mm}^2$ reticle is repeated to pattern the entire wafer. Random effects are averaged over the nearly one hundred stages of the oscillator and are hardly visible.

These large distance effects do not have a direct impact on equality of transistor pairs. Most of this effect is a global variation and becomes part of the tolerance budget in the definition of the parameter corners of a process. In some digital design environment a specific guard band is used to make sure that the potential loss of current drive capability becomes visible to the designer. Yet the example makes clear what the significance is of lithography variations on the overall performance.

5.2.5 *Proximity Effects*

Figure 5.12 visualizes proximity effects on a group of lines. If the proximity effect is caused by the diffused light from neighboring fields, the line width in the open field will become narrower. Large neighboring structures cause lines to expand. In a precision lay-out dummy structures are placed at distances up to 20–40 µm. Proximity effects can be caused not only by lithography, but also by depletion of etch liquids or gases.

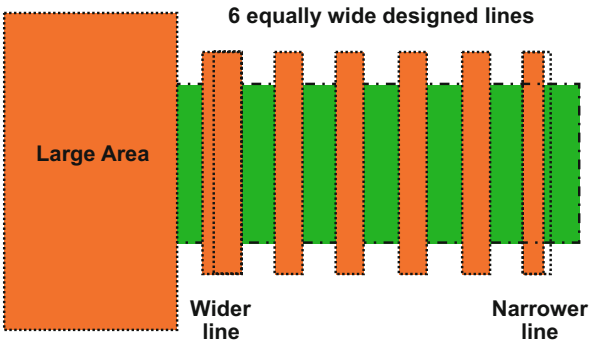


Fig. 5.12 The proximity effect: lines with large neighboring structures grow in size, while lines next to open space shrink

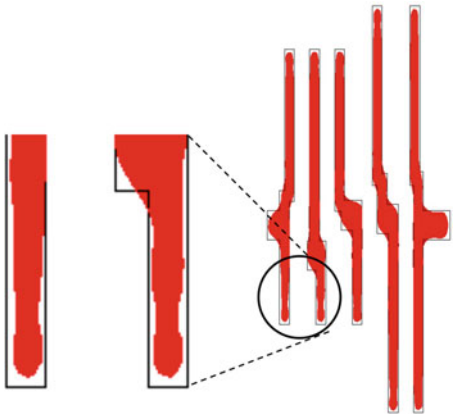


Fig. 5.13 The tips of the structures are enhanced by the OPC tool in this lithography simulation

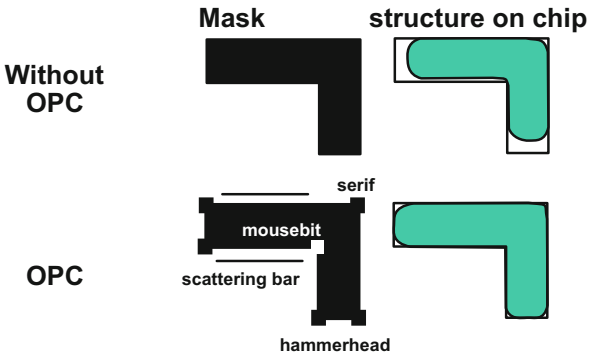


Fig. 5.14 Standard lithography results in rounded structures that are too short. With the help of optical proximity corrections in the mask the net result on a wafer comes much closer to the desired dimensions

An example of dimensional deformations of an advanced lithographic tool is visible in the lithography simulation of Fig. 5.13.

Patterns in one layer may sometimes affect the patterning in other layers. During the spinning of the resist fluid, resist may accumulate against altitude differences of previous layers on a partly processed wafer. This results in circular gradients and is therefore often not easily recognized as a systematic offset.

In 65-nm technologies it is practically impossible to define minimum width lines with acceptable tolerance at random positions. In order to create minimum width patterns pre-distortion is applied to the mask in the form of optical proximity correction (OPC), Fig. 5.14.

5.2.6 *Implantation Related Effects*

An ion beam from an implanter that is perfectly aligned with the crystallographic orientation of the substrate will result in ions to penetrate deeply into the lattice. This effect is called "channeling." During the ion-implantation steps in older processes the implantation beam is tilted by some 5–8°. As a result of this non-perpendicular implantation, channeling is avoided but source and drain diffusions will be asymmetrical. The diffusion on one side may extend further underneath the gate than on the other side, see Fig. 5.15. In order to prevent inequalities in currents or overlap capacitors, the directions in which the MOS currents flow must be chosen to run parallel, and not rotated or anti-parallel. In integrated circuit manufacturing there are more processing steps that can cause similar asymmetries: e.g., mask alignment deviation of the contact hole mask in Fig. 5.15 (left) will create a positive change in source series resistance on one side of a mirrored structure $+\Delta R_S$ while the other side sees the opposite shift $-\Delta R_S$.

The well-proximity effect in Fig. 5.16 has no direct relation with lithography. This effect is believed to be caused during the implantation of the well in the substrate. The implanted ions interact with the photo-resist boundary and cause a horizontal gradient in the well implantation dose. Variations ranging from 1 μm [65] to 2 μm [66] have been reported. Although modern process flows use much steeper

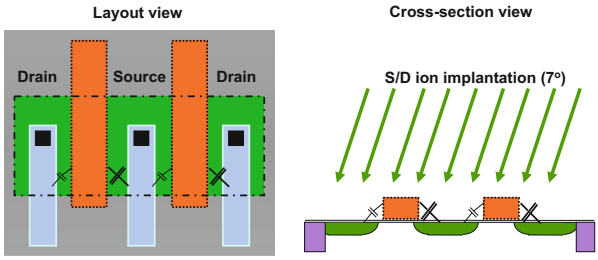


Fig. 5.15 The source and drain diffusions are implanted under an angle. This causes asymmetry for drain and sources

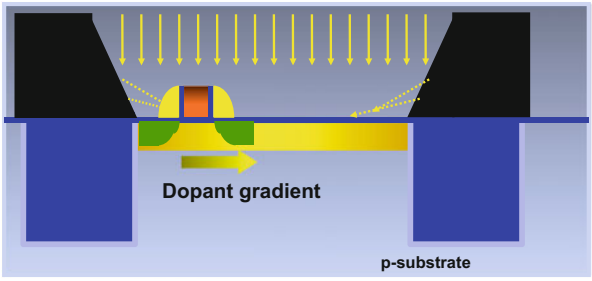


Fig. 5.16 The well-proximity effect occurs during the well implant. The drawn transistor is fabricated afterwards, but resides in a well with a horizontal doping gradient [65, 66]

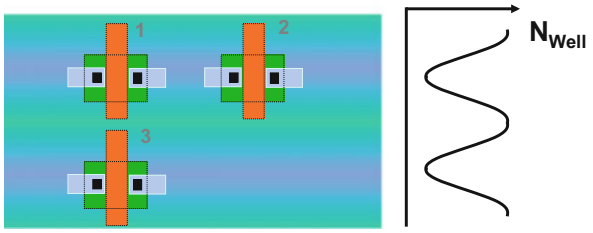


Fig. 5.17 A large area is implanted by scanning the implanter beam over the wafer area. A sharply focussed beam will cause stripes and a doping profile in a direction perpendicular to the scanning direction. Differences in threshold voltages will occur

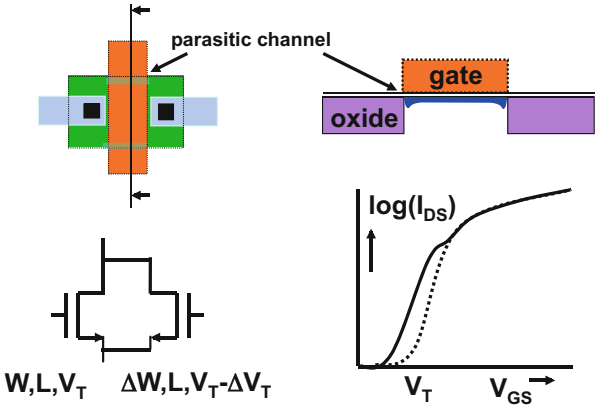


Fig. 5.18 The effect of parasitic channels along the field oxide

photo-resist boundaries, it is advisable to use an overlap beyond the minimum lay-out design rule for implantation masks, where possible.

Many effects can occur in a process that affect the reproducibility of transistors and other components. Figure 5.17 shows three transistors in an n-well. A larger structure like a well is implanted by scanning the implantation beam stripe by stripe over the area. A sharply focussed beam will create a distinguishable stripping pattern with a doping gradient running perpendicular to the scanning direction. As a consequence transistors 1 and 2 will match, but transistor 3 will deviate.

A lot of attention is normally focussed on the channel to source and drain perimeter. The boundary of the channel to the surrounding isolation material, such as field oxide or shallow trench layers, is equally important. In Fig. 5.18 the doping profile next to the isolating oxide is such that a parasitic channel can occur. This can happen when a well doping and the substrate doping locally compensate each other. The effect is that the desired transistor is flanked by two narrow width, but low-threshold voltage transistors. Especially in low current regime and in weak inversion

the parallel transistors will generate a considerable current, often indicated by a bump in the sub-threshold $I_{DS} - V_{GS}$ characteristics. Slight differences in biasing between transistors of a matched pair lead to considerable current deviations.

5.2.7 Temperature Gradients

Gradients can exist in doping, resistivity, and layer thickness. Although structures tend to decrease in dimensions, situations may occur in which equality is required in a distance in the order of 1 mm. In older processes CMOS thresholds were observed to deviate up to 5 mV over this distance.[4] Resistivity gradients can reach a relative error of several percent over this distance. In advanced processes ($<0.18\,\mu m$) process control is much better and technology gradients are hardly present.

The temperature distribution across a circuit in operation can be a reason for parameter gradients, see Fig. 5.19. In a System-on-Chip the different blocks show a great variety of power dissipations. On-chip memories shows a relatively low power density. Output drivers, transmitters, high-speed processors, power regulators (LDO), and input stages (LNA) may consume much more. In larger chips (50–100 mm^2) with 2–5 W power dissipation temperature differences up to 20 °C can occur. Local temperature gradients of 2–5 °C/mm are possible. With threshold-voltage and diode temperature coefficients of -2 mV/°C, an offset in the order of several mV is well possible. It is therefore important to consider the power distribution when temperature sensitive circuits and heat sources are placed on the

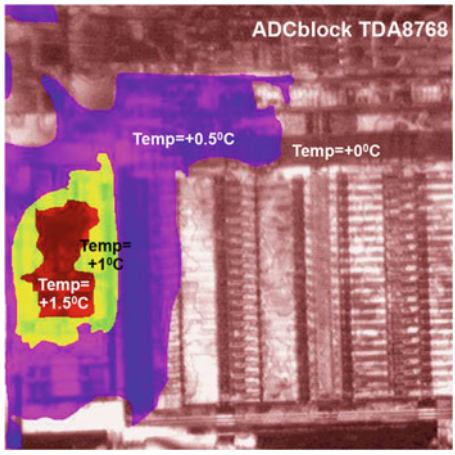


Fig. 5.19 Temperature difference can be visualized by means of a liquid crystal technique. Here the track-and-hold circuit generates the heat, which spreads out via the wiring into the ladders of this analog-to-digital converter

[4] A rough estimate for a technological gradient can be approximated by dividing the difference between the maximum and minimum of the parameter by half of the wafer diameter.

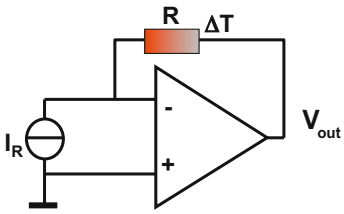


Fig. 5.20 A high-quality sinusoidal current is converted into a distorted output voltage because the resistor will heat up

same die. The circuit designer can position the critical devices on equal-temperature lines or use cross-coupling.

Example 5.3. A current source $I_R(t)$ delivers a perfect sinusoidal current. This current is converted into a voltage by means of a high gain high bandwidth opamp and an on-chip resistor, Fig. 5.20. The resistor has a negligible voltage coefficient and a temperature coefficient κ. After processing, a severe distortion appears at the output for low frequency signals. What is a potential reason for this distortion?

Solution. During a sine wave the current through the resistor varies and consequently the momentary heating. As the thermal relaxation of IC components is in the range of microseconds, the temperature of the resistor will follow the magnitude of the dissipated power. At any moment the temperature rise will be proportional to the dissipated power and the thermal conductivity: $T - 25°C = I_R^2(t)R\kappa$. So the effective resistance is: $R(T) = R(T = 25\,°\text{C})(1 + TC \times (T - 25\,°\text{C}))$.

So the output voltage now is: $V_{out} = I_R(t)R(25\,°\text{C})(1 + TC\kappa I_R^2(t)R(25\,°\text{C}))$. In case of a sine wave with zero bias, a third order distortion component will appear. The magnitude of the distortion will reduce at higher signal frequencies.

5.2.8 *Offset Caused by Stress*

During the fabrication of a circuit, layers are deposited in and on the substrate. These layers are built of different materials with different thermal expansion coefficients. After the devices have been cooled to room temperature differences in thermal expansion coefficients lead to mechanical stress. This stress can result in a positive or negative change of the local parameter value. A secondary effect is that the global stress pattern is locally affected by neighboring lay-out features, causing stress modulation in the surrounding components. In high-precision analog design this will lead to undesired systematic offsets.

In resistors and transistors stress predominantly impacts the mobility of carriers. Tensile stress increases the electron mobility and reduces the hole mobility. Compressive stress works opposite and is used to enhance mobility in PMOS transistors. An effect on threshold voltage does occur as well, see Fig. 5.23. Some major causes for stress are

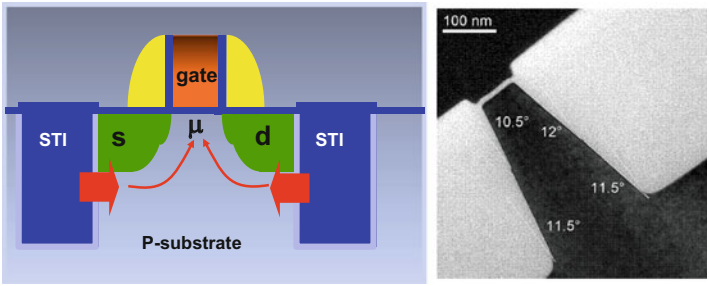


Fig. 5.21 Stress in integrated circuits is caused by thermal expansion of various materials. The blocks of shallow trench isolation (STI) create a considerable stress in the active area of the transistor, causing current variations up to 10 %. The photo shows a cross-section, where the slopes of the STI are indicated (Courtesy: C. Detcheverry, NXP)

- In modern isolation techniques a trench is etched in the substrate and silicon dioxide is deposited and planarized: shallow trench isolation (STI), Fig. 5.21. The different thermal coefficients of silicon dioxide and the substrate cause mechanical stress. A transistor in the substrate with its diffusion areas is surrounded by STI and experiences this stress. This effect is called "STI-stress" or "LOD-stress" (Length Of Diffusion).
- In a densely packed circuit, there will be many active-to-STI edges. Unrelated edges that are close to a device will influence the stress pattern, this effect is known as "OD-to-OD" stress or "OD-spacing" effect. The oxide-definition (OD) mask defines the inverse of the active area.
- In advanced processes an etch-stop layer can be used to create stress that increases the current in a high-performance MOS transistor. The proximity of other structures will influence this effect, called "PS-to-PS" stress or "poly space" effect.
- Aluminum has a different thermal expansion coefficient from the dielectric that surrounds it. Asymmetries in wiring will result in stress related offset.
- The presence of the die boundary close to sensitive devices. Typically a distance of several hundreds of microns is safe to avoid die-edge related stress effects.
- Plastic packages are molded around the die. After cooling these packages create rather severe mechanical stress. Special gels or polyimide coatings on top of the die relieve this problem. A simple way to detect package related stress is to heat the package with a hot air flow.

Stress related to shallow trench isolation has been subject of various studies [66–69]. At temperatures of over 1000 °C areas of silicon dioxide are formed in the substrate. At that temperature the structure is free of stress or relaxed. After cooling the difference in thermal expansion coefficient causes a compressive mechanical stress that peaks at the border of the active area and the STI formation [67]. The mechanical stress deforms the lattice and causes the mobility in the transistors to vary.

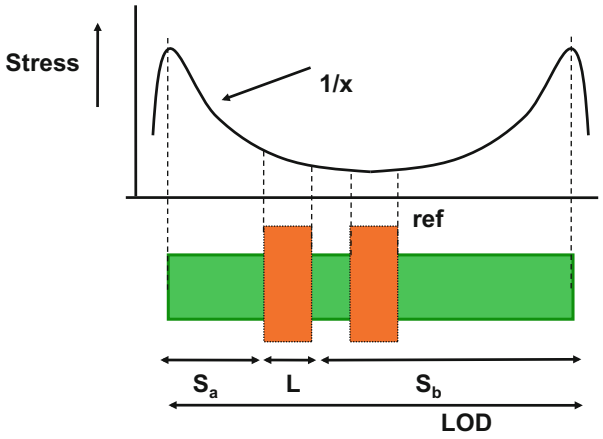


Fig. 5.22 A transistor is placed asymmetrically with respect to the reference device. The edges between active and STI inflict stress in the channel region, however, due to the asymmetry the effect of the stress is different for both devices

Figure 5.22 shows two transistors that are placed asymmetrically with respect to the edges of the active area and the STI isolation. The mobility of the device μ_{eff} is usually modeled with an inverse distance model [67]:

$$\begin{aligned} S &= \frac{1}{S_a + L/2} + \frac{1}{S_b + L/2} \\ \mu_{eff} &= \frac{1 + K_{s,\mu} S}{1 + K_{s,\mu} S_{ref}} \mu_{ref} \\ V_T &= V_{T,ref} + K_{s,VT}(S - S_{ref}) \end{aligned} \tag{5.2}$$

where the suffix "*ref*" indicates the reference device and K_s is a process constant. The parameter S reflects the distance parameter. The stress from the STI edges affects also the doping profile under the transistor. This phenomenon is not fully understood, however, the idea that a lattice deformation changes the diffusion of the doping atoms, seems plausible.

In [69] an experiment is reported where the STI-to-active edge of the source and drain is varied. Figure 5.23 shows current factor deviations up to 12 % and threshold-voltage variations of 10 mV. These observations are technology specific but similar effects are reported in [66, 67].

Next to the effect of the STI-to-active edge of the transistor itself, also neighboring edges will modulate the mechanical stress pattern. This may be less relevant in a digital circuit, however, in precision analog design these effects must be taken into account. The generalized model uses:

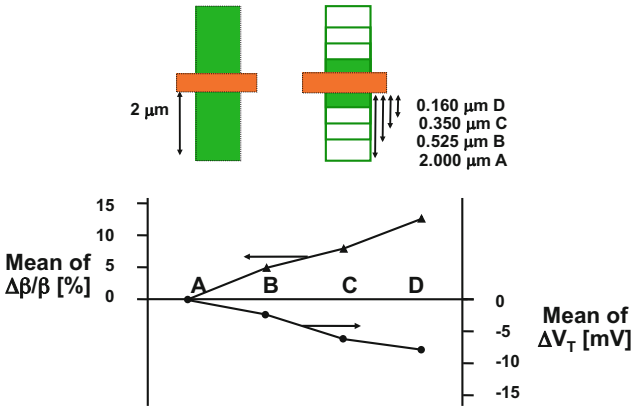


Fig. 5.23 An experiment shows the influence of the STI edge on the drain current and threshold voltage [69]. *Top*: a 65-nm 2/0.5 µm NMOS reference transistor is designed with the STI edge of the source and drain at 2.0 µm. A second device has a similar STI distance (A) or at 0.525 (B), 0.35 (C), and 0.16 µm (D)

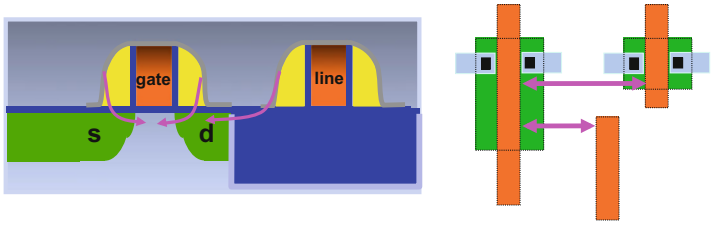


Fig. 5.24 The poly space effect is caused by the etch-stop layer

$$S = \sum_{i=1}^{n} \frac{\pm W_i}{2W} \frac{1}{S_i + L/2} \tag{5.3}$$

where the first term relates the width of the edge to the gate width and the ± sign to an STI-active edge or an active-STI edge.

The poly space effect [70] in Fig. 5.24 is caused by the stress related to the use of the etch-stop layer. This layer applies compressive or tensile stress to the transistor in order to increase the current drive capabilities. Also stress coming from neighboring devices will influence the stress pattern under the critical transistors. Again an inverse distance model applies.

5.2.9 Issues with Wiring

The wiring pattern causes transistors to show offset. The coverage of transistors with metal layers can lead to mobility reduction due to incomplete annealing of interface states [71] and to variations in the stress pattern. When a wiring pattern is placed at different spacings on one side of a bipolar pair [72], the resulting current variation is in the order of 1 %.

Lithography critically depends on the flatness of the wafer surface. Damascene wiring technology has been developed to avoid height differences. In a design with non-regular wiring patterns, the design rules will prescribe filling patterns: "tiling". If the design tool is allowed to automatically generate tiling patterns, undesired side effects may occur. It is clear that the proximity of a tiling structure will affect capacitor ratios. Also stress patterns and thickness variations can occur. A safe approach is to define and position the tiling patterns during the lay-out phase by hand. Also wiring on top level (e.g., tiling patterns) can cause stress [73]. Copper wiring behaves similar to aluminum wiring: deviations in the 1–2 % range.

Figure 5.25 suggests that the impact of the wiring pattern halves for every 10 µm distance up to 40 µm. This example again shows that a regular, symmetrical, and consistent lay-out is required for analog circuits that should yield offsets below 1 %. Table 5.4 lists a number of effects.

5.2.10 Offset Mitigation

Systematic deviations are mostly identified during the initial design trials of a process. Sometimes optimizations in an established process flow still lead to unexpected deviations. Measures for overcoming them are found in an extensive study of the fabrication process. As dimensions shrink, some offset effects are incorporated in the device model descriptions that quantify the impact [12]. Despite the complex nature of some variations, a number of guidelines can be formulated to minimize the effect of these offset causes, see Table 5.5 [74].

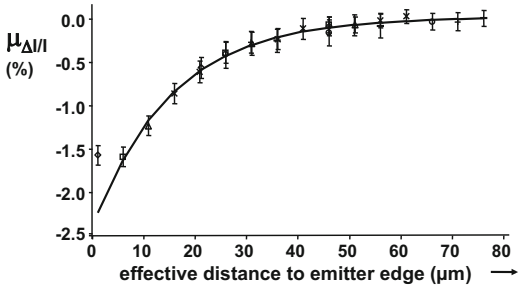


Fig. 5.25 An aluminum wire placed at a certain distance of the emitter causes current deviations, that can be measured up to 40 µm [72]

Table 5.4 Potential magnitudes of deterministic variability effects

Effect	Magnitude
Power supply voltage drop	Voltage shifts up to 100 mV in poorly designed power grids.
Lithography, etch depletion, etc.	Dimensional errors up to 100 nm, in >0.25 µm CMOS
Proximity effects	Dimensions errors up to 20 nm in advanced mask making
Temperature gradient	1 °C over 20–50 µm close to strong local heat source. With 2 mV/°C threshold sensitivity
LOD or STI effect [66, 67, 69]	Mobility changes 12 % between minimum drain extension and large drain
Well-proximity [65, 66]	Threshold shift up to 50 mV if mask edge closer than 2 µm
Metallization [72]	Mobility changes 2 % between close and 40 µm far metal track
Metal coverage of gates [71]	Currents may deviate up to 10–20 %

Table 5.5 Guidelines for the design of matching components

1	Matching components are of the same material, have the same form, dimensions, and orientation
2	The potentials, temperatures, pressures, and other environmental factors are identical
3	Currents in components run in parallel, not anti-parallel, or perpendicular
4	Only use cross-coupled structures if there is a clear reason for that (e.g., temperature gradient). Identify the heat centers
5	Avoid overlay of wiring on matching components.
6	Use star-connected wiring for power, clock, and signal
7	Apply symmetrical (dummy) structures up to 20 µm away from sensitive structures
8	Keep supply and ground wiring together and take care that no other circuits dump their return current in a ground line
9	Check on voltage drops in power lines
10	Stay 200 µm away from the die edges to reduce stress from packaging
11	Tiling patterns are automatically inserted and can lead to unpredictable coupling, isolation thickness variations, and stress. Do not switch off the tiling pattern generation, but define a symmetrically placed tiling pattern yourself

Common centroid structures are used to reduce the gradient effects [75]. Applying a common centroid geometry is not trivial, an asymmetry in the wiring scheme can easily cause more problems than are solved, see Fig. 5.26. On the left side is a standard cross-coupled differential pair with common source. The right side shows in-line common centroid structures. The lower structure is exactly common centroid with the disadvantage that the outer devices need dummy structures to compensate for their lack of neighbors. The upper-right structure needs no dummy structures, at the cost of a small spacing between the common centroid points.

By following these design guidelines the effects of systematic errors can be significantly reduced. The obtainable limits in a production environment differ per component and can be summarized as:

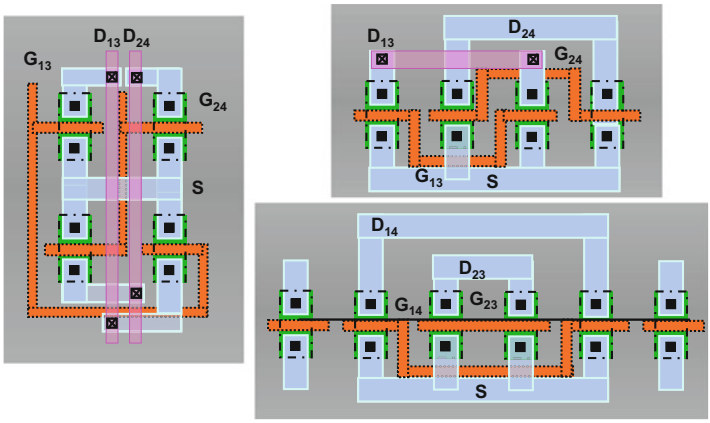


Fig. 5.26 Some common centroid arrangements

- Resistors: The absolute value suffers from process variations and temperature. Yet, the relative accuracy of matched resistors is in the order of 10^{-3}–10^{-4} depending on type (diffused is better than polysilicon), size, and environment. Large resistor structures are sensitive to substrate coupling.
- Capacitors: The absolute value is usually well-defined in a double polysilicon or MIM process. Also horizontally arranged capacitors such as fringe capacitors reach excellent performance. The relative accuracy of capacitors is in the order of 10^{-4} for $>$ 1 pF sizes. Minimum usable sizes in design are limited by parasitic elements, relative accuracy or the kT/C noise floor. In the application the net effect of the capacitor is sensitive to different parasitic couplings, which can be mitigated with stray-capacitor insensitive circuit topologies. Often capacitors are seen as a low power solution, but handling charges requires large peak currents during transfers, so the power of the surrounding circuits limits the low power ambition of capacitor based circuit solutions.
- Transistors: The current is sensitive to temperature, process spread, and variability effects. The relative accuracy in current is in the order of 10^{-3}. Back-gate modulation by substrate noise and 1/f noise must be considered.
- Time: with a more or less fixed timing variation or jitter (1–5 ps_{rms}), the best accuracy is achieved for low signal bandwidths.

Example 5.4. In Fig. 5.27 a cross-coupled transistor lay-out is shown. What is the most important mistake in the lay-out?

Solution. There are several minor mistakes in this lay-out: The upper-right transistor is partially covered by metal-2(purple). This will create stress and potentially annealing problems. The connections of the drains in the middle are far from symmetrical. The current will flow from top side to bottom side and become part of a current loop. However, the most important mistake is the placement of the joint

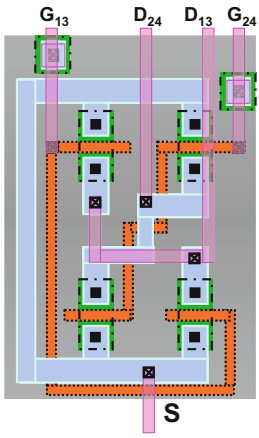


Fig. 5.27 A common centroid arrangement

source "S" connection. In the connection from the lower left transistor to the "S" terminal now flow the currents of three transistors, instead of the current of a single transistor. It is better to connect the "S" terminal in the middle of the left side metal-1 connection.

5.3 Random Matching

5.3.1 Random Fluctuations in Devices

Parameters that reflect the behavior of devices are the result of the combination of a large number of microscopic processes. The conductivity of resistors and transistors and the capacitance of capacitors is built up of a large number of single events: e.g., the presence of ions in the conduction path, the local distances between the polysilicon grains that form capacitor plates, etc. Already in 1961 W. Shockley recognized that these atomic processes can lead to random fluctuations of device parameters [76]. Various authors have investigated random effects in specific structures: capacitors [77–82], resistors [83, 84], MOS transistors [85–89], and bipolar devices [90].

In a general parameter fluctuation model [91] a parameter P describes some physical property of a device. P is composed of a deterministic and random varying function resulting in varying values of P at different coordinate pairs (x, y) on the wafer. The average value of the parameter over any area is given by the weighted integral of $P(x, y)$ over this area. The actual difference between two parameters P of two identically sized areas at coordinates (x_1, y_1) and (x_2, y_2) is

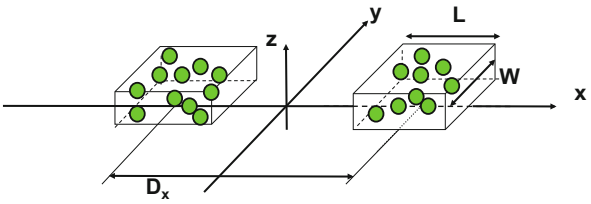


Fig. 5.28 Definition of the area function $h(x, y)$

$$\Delta P(x_{12}, y_{12}) = P(x_1, y_1) - P(x_2, y_2) =$$

$$\frac{1}{\text{area}}\left[\int\int_{\text{area}(x1,y1)} P(x', y')dx'dy' - \int\int_{\text{area}(x2,y2)} P(x', y')dx'dy'\right] \tag{5.4}$$

This integral can be interpreted as the convolution of double box functions formed by the integral boundaries (or the device dimensions) with the "mismatch source" function $P(x, y)$. In the Fourier domain the convolution transforms in a multiplication and allows separating the geometry-dependent part from the mismatch source:

$$\Delta\mathcal{P}(\omega_x, \omega_y) = \mathcal{G}(\omega_x, \omega_y)\mathcal{P}(\omega_x, \omega_y) \tag{5.5}$$

Now the mismatch generating process $\mathcal{P}(\omega_x, \omega_y)$ can be regarded as a source that generates spatial frequencies that are spatially filtered by the device geometry dependence function $\mathcal{G}(\omega_x, \omega_y)$. These two components are analyzed separately.

The geometry function as shown in Fig. 5.28 for a pair of rectangular devices with area WL is defined as:

$$h(x, y) = \begin{cases} \frac{1}{WL}, & (D_x/2 - L/2) < x < (D_x/2 + L/2), -W/2 < y < W/2 \\ \frac{-1}{WL}, & (-D_x/2 - L/2) < x < (-D_x/2 + L/2), -W/2 < y < W/2 \\ 0, & \text{elsewhere} \end{cases} \tag{5.6}$$

For convenience it has been assumed that both areas are at a distance D_x along the x-axis. Some mathematical manipulation results in a geometry function for the difference in **paired** transistors parameters, Fig. 5.28:

$$\mathcal{G}(\omega_x, \omega_y) = \frac{\sin(\omega_x L/2)}{\omega_x L/2}\frac{\sin(\omega_y W/2)}{\omega_y W/2}[2\sin(\omega_x D_x/2)] \tag{5.7}$$

This geometry function has a zero value for $\omega_x = 0$, thereby eliminating the global value of the parameter from the calculations. The geometry functions for other geometries are found in the same way, e.g., a cross-coupled group of four transistors as in Fig. 5.26 (left) has a geometry function where the last term in brackets in Eq. 5.7 is replaced by $[\cos(\omega_x D_x/2) - \cos(\omega_y D_y/2)]$.

After this analysis of the geometry dependence the specification of the random contribution to $P(x, y)$ or $\mathcal{P}(\omega_x, \omega_y)$ has to be formulated.

Different classes of distinct physical mismatch causes can be considered to describe local and global variations. Every mismatch generating physical process that fulfills the mathematical properties of such a classes results in a similar behavior at the level of mismatching transistor parameters.

The most relevant class of phenomena is characterized by a random process on a parameter P if:

- The total mismatch of parameter P is composed of mutually independent events of the mismatch generating process.
- The effects on the parameter are so small that the contributions to the parameter are linear.
- The correlation distance between the events is small compared to the size of the device (basically saying that boundary effects can be ignored).

In the frequency domain this type of spatial random processes is modeled as spatial "white noise." A process with these properties is described in the Fourier domain as a constant value for all spatial frequencies.

The combination of this mismatch generating process and the paired transistor geometry function results in a description of the power or the variance of the difference ΔP in parameter P between the two instances [91]:

$$\sigma^2_{\Delta P} = \frac{A_P^2}{WL} \tag{5.8}$$

A_P is the area proportionality constant for parameter ΔP. The proportionality constant can be measured and used to predict the mismatch variance of a circuit.

Many known processes that cause mismatching parameters fulfill in first order the above mentioned mathematical constraints: distribution of ion-implanted, diffused or substrate ions (random dopant fluctuations), local mobility fluctuations, polysilicon and oxide granularity, oxide charges, etc.

Equation 5.8 describes the statistical properties of area *averaged* or *relative* values of parameter P. The *absolute* number of events (like the charge in an MOS channel) is proportional to the area of the device WL. Therefore differences in the sums of atomic effects obey a Gaussian distribution with zero mean and

$$\sigma_{\Delta P} = A_P \sqrt{WL} \tag{5.9}$$

In analyzing statistical effects it is important to consider whether the parameter is an absolute quantity (e.g., total amount of ions) or is relative (averaged) to the device area (e.g., threshold voltage).

Apart from theoretical derivations and measurements, 3-D device simulations are applied to analyze the impact of random dopants, line-edge roughness, and polysilicon granularity in advanced processes [92].

The assumption of a short correlation distance in the above process implies that no relation exists between matching and the distance D_x between two transistors. Wafer maps show, however, all sorts of parameter-value distributions that originate from wafer fabrication. This effect is observed in Fig. 5.11. The extremes of a parameter are often located in the middle and at the edges of a wafer. A rough approximation is obtained by dividing the maximum process spread, e.g., 100 mV for a threshold spread or 10 % for the current factor by half of the wafer diameter. Gradients in the order of 1 mV/mm result and are small compared to random mismatch.

5.3.2 *Probability Density Function*

Equation 5.8 indicates the variance of the mismatch process as obtained from a variance analysis in the spatial frequency domain. The probability distribution function for various random process that obeys the previously described boundary conditions, still has to be established.[5] For implantation and diffusion processes Shockley notices in his 1961 paper on breakdown of p-n junctions [76, p. 48]: "Thus in the following we shall assume that the chemical charges of donors and acceptors are distributed statistically and independently in accordance with a Poisson distribution."

The Poisson distribution, Fig. 5.29 is applied in cases where the average outcome λ of a statistical process is known. The probability that the actual outcome k will occur is

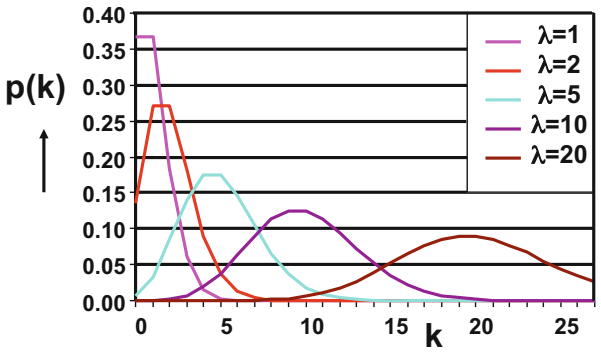


Fig. 5.29 The Poisson distribution for various values of the average

[5]Compare thermal noise: a flat spectrum in the frequency domain and a Gaussian amplitude distribution.

$$p(k) = \frac{e^{-\lambda}\lambda^k}{k!} \tag{5.10}$$

The Poisson probability density distribution has the important property:

$$\sigma^2_{Poisson} = \mu_{Poisson} = \lambda$$

The variance equals the mean value and both equal the average parameter.

For large values of λ the Poisson distribution converges to a Gaussian distribution, but keeps the properties of a Poisson distribution:

$$\sigma^2_{Gauss} = \mu_{Gauss} = \sigma^2_{Poisson} = \mu_{Poisson} = \lambda, \text{ for } \lambda \to \infty$$

Thereby the mean and the variance of the resulting Gaussian probability density function are the same.

The distribution of charges in a depletion region, a resistive structure, a base of a resistor, the dielectric of a capacitor, and more structures are subject to this form of analysis.

When very small structures are analyzed where the average number of relevant dopant atoms is low, or phenomena with inherently averages, e.g., $1/f$ noise centers, the Poisson distribution is used, for most other cases λ can be supposed to be sufficiently large to use the Gaussian distribution.

The values of parameter P in the previous paragraph are also described by a Poisson process that converges for a large number of events to a Gaussian distribution. The difference between the parameters P of two structures ΔP is now a Gaussian distribution with

$$\mu_{\Delta P} = 0, \qquad \sigma^2_{\Delta P} = 2\lambda$$

5.3.3 *Functions of Statistical Variables*

In device physics and in circuit design the statistical phenomena occur at a very basic level, e.g., the dopant distribution. Often a designer requires to translate this basic behavior into properties of the circuit, or even chip. The mathematical toolbox to propagate statistical parameters starts with the elementary function $y = f(x)$. This function describes, e.g., the transfer function of device parameters into currents or voltages. If y relates to x via a smooth and differentiable function, as in Fig. 5.30, the mean and variance of y can be approximated by the partial derivative using the Taylor series expansion [9]:

$$\mu_y = E(f(x)) \approx f(\mu_x) + \left(\frac{d^2 f(x)}{dx^2}\right)\frac{\mathrm{Var}(x)}{2} = f(\mu_x) + \left(\frac{d^2 f(x)}{dx^2}\right)\frac{\sigma_x^2}{2}$$

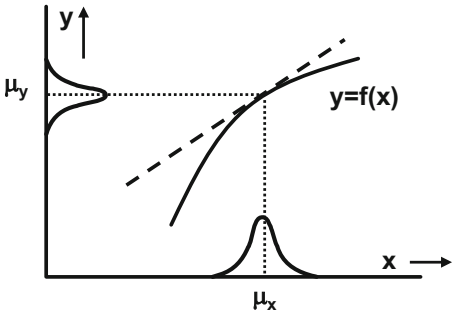


Fig. 5.30 The transformation of one stochastic variable in another via a function $y = f(x)$ uses the Taylor expansion

$$\sigma_y^2 = E(f^2(x)) - (E(f(x)))^2 \approx \left(\frac{df(x)}{dx}\right)^2 \mathrm{Var}(x) = \left(\frac{df(x)}{dx}\right)^2 \sigma_x^2$$

Equivalently the propagation of random variables x_1 and x_2 in the relation $y = g(x_1, x_2)$ can be calculated:

$$\sigma_y^2 \approx \left(\frac{dg(x_1,x_2)}{dx_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{dg(x_1,x_2)}{dx_2}\right)^2 \sigma_{x_2}^2 + \left(\frac{dg(x_1,x_2)}{dx_1}\right)\left(\frac{dg(x_1,x_2)}{dx_2}\right) \mathrm{Cov}(x_1,x_2)$$

For independent variables the covariance $\mathrm{Cov}(x_1, x_2) = 0$ and this result reduces to:

$$\sigma_y^2 \approx \left(\frac{\partial g(x_1)}{\partial x_1}\right)^2 \sigma_{x1}^2 + \left(\frac{\partial g(x_2)}{\partial x_2}\right)^2 \sigma_{x2}^2 \tag{5.11}$$

The above equation can be easily expanded to three or more input variables. If the function g is a simple sum of terms, the variance of g equals the well-known sum of the variances of its composing terms, e.g., for the sum or differences of normal distributed variables :

$$\begin{aligned} y = g(x_1, x_2, x_3, \ldots) &= a_1 x_1 + a_2 x_2 + a_3 x_3 \ldots . \\ E(y) &= a_1 E(x_1) + a_2 E(x_2) + a_3 E(x_3) \ldots . \\ \sigma_y^2 &= a_1^2 \sigma_{x1}^2 + a_2^2 \sigma_{x2}^2 + a_3^2 \sigma_{x3}^2 \ldots \end{aligned} \tag{5.12}$$

The squaring operation on the partial derivatives of the variance causes that the variance cannot have a "minus" sign in front of the constituent terms. In other words: every addition increases the variance.

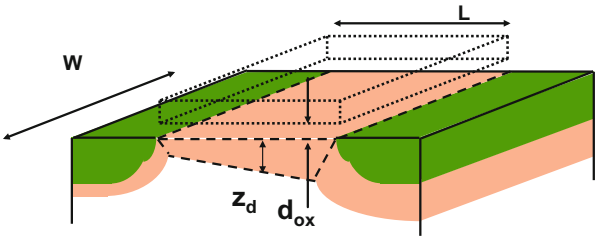


Fig. 5.31 A cross-section through an MOS transistor indicating the depletion region

5.3.4 MOS Threshold Mismatch

The threshold voltage is given by:

$$V_T - V_{FB} = \frac{Q_B}{C_{ox}} = \frac{qN_x z_d}{C_{ox}} = \frac{\sqrt{2q\epsilon N_x \phi_b}}{C_{ox}} \tag{5.13}$$

where ϵ is the permittivity, N_x the dope concentration, and ϕ_b the Fermi potential. If the depletion region of a transistor (see Fig. 5.31) is defined by its width W, length L, and a depletion region depth $z_d = \sqrt{2\epsilon\phi_b/qN_x}$, then the volume of the depletion region is (in first order): WLz_d. Different impurities are active in this region, with concentrations around 10^{16}–10^{18} cm^{-3}. N_x contains acceptor and donor ions from the intrinsic substrate dope, the well, threshold adjust, punch-through implantation.[6] In the variance analysis it is important to note that the total number of charged ions and other charge contributions must be considered, not the net resulting charge.

The variance in the number of ions is now approximated by Poisson statistics:

$$\sigma^2 = \mu \quad \mu = WLz_d N_x, \quad \Rightarrow \sigma_{WLz_d N_x} = \sqrt{WLz_d N_x} \tag{5.14}$$

The threshold variance can now be derived from Eq. (5.13) by considering that the variance of a threshold voltage equals the variance of the charge in the depletion region multiplied by the partial derivative of the threshold versus the charge. With Eq. 5.11 the variance for a single threshold voltage due to the fluctuation in the number of dopant is obtained:

$$\sigma_{VT}^2 = \left(\frac{\partial(V_T)}{\partial(WLz_d N_x)} \right)^2 \sigma_{WLz_d N_x}^2 \tag{5.15}$$

As matching usually occurs between pairs of transistors, the variance of the *difference between two transistors*, each with an area WL, is [91]:

[6] For ease of understanding only a uniformly distributed dopant is assumed, more complicated distributions must be numerically evaluated.

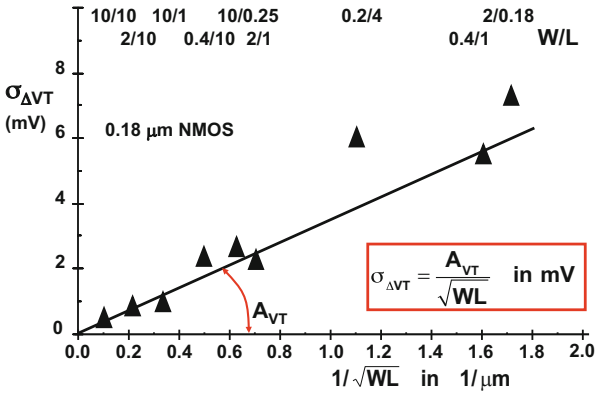


Fig. 5.32 The standard deviation of the NMOS threshold and the relative current factor versus the inverse square root of the area, for a 0.18 μm CMOS process

$$\sigma_{\Delta VT} = \sqrt{2}\sigma_{VT} = \frac{qd_{ox}\sqrt{2N_x z_d}}{\epsilon_{ox}\sqrt{WL}} = \frac{A_{VT}}{\sqrt{WL}} \tag{5.16}$$

This function is commonly depicted as a linear relation between $\sigma_{\Delta VT}$ and $1/\sqrt{\text{area}}$. Figure 5.32 shows an example of the measured dependence for $\sigma_{\Delta VT}$ versus $1/\sqrt{\text{area}}$. The slope of the line equals the parameter A_{VT}. In this plot the lay-out dimensions were used. For the smallest sizes the effective gate area is smaller due to under-diffusion and channel encroachment.

The basic threshold-voltage mismatch model[7] has been extended by various authors. Forti and Wright [88] showed that mismatch parameters obtained in the strong inversion regime are also applicable in the weak inversion regime. More geometry dependence factors can be included to address deep-sub-micron effects [93]. A correction for the vertical distribution of charges was derived in [94]. Work reported in [69] indicates that there is no relation between deterministic variations and random dopant fluctuations.

Andricciola and Tuinhout [95] shows that the threshold-voltage mismatch and the relative current factor mismatch are hardly affected by temperature.

In deep sub-micron processes the short channel effects in the example channel are controlled by means of "halo" or "pocket" implants. These implants are self-aligned with the gate stack and can introduce some significant variations in the local doping profiles. Next to their own variation, the self-aligned feature prints any line-edge roughness in the doping profile. The pocket implants defy the uniform dopant

[7] An often proposed "happy-hour" mismatch model explains threshold fluctuation as a form of petrified $1/f$ noise. Threshold mismatch is determined by the number of charged ions and even for the smallest size transistors there are still on average >100 ions. The same transistor shows none or just one or two electron trapping centers. The numbers don't fit. The trapping centers also follow a Poisson distribution and follow an area scaling law.

hypothesis for the calculation of the threshold mismatch of Eq. 5.8. An additional term can be included for the variation due to the pocket implant:

$$\sigma^2_{\Delta VT} = \frac{A^2_{VT}}{WL} + \frac{B^2_{VT}}{f(W,L)} \tag{5.17}$$

where the function $f(W, L)$ of the width and length still needs to be established [96].

Technologies with feature sizes below 40 nm mostly use metal-gate constructions. These gates have several advantages over polysilicon gates: no depletion of the polysilicon grains and an easy setting of the threshold voltage by means of the work function (WF) of the metal. The work function is the energy in electron-Volt needed to pull out an electron from its atom and move it into vacuum. This work function is part of the threshold voltage and appears in Eq. 5.13 in the factor V_{FB}. The metal for constructing the transistor gate consists of titanium-nitride or alloys additionally containing aluminum and carbon. When deposited, grains are formed of a few nanometer size. The work function of a grain varies with the orientation with respect to the lattice. The work function is anisotropic, Fig. 5.33. So, the overall work function is a random distribution of many grain oriented in various directions with different work functions. With sufficient grains forming the transistor gate, the distribution of the effective work function in the threshold voltage approximates a gauss probability density and the above math can be applied, with the same mathematical result for its variation as in Eq. 5.16.

Example 5.5. In Fig. 5.34 two highly doped n-regions form diodes in a low-doped substrate $N_A = 10^{16}\,\mathrm{cm}^{-3}$. Each diode area is $1\,\mu\mathrm{m}^2$.

Solution. The depletion width is found from: $z_D = \sqrt{2\epsilon_r\epsilon_0 V_{pn}/qN_A}$. The result is $z_D = 0.3\,\mu\mathrm{m}$. The diode current is

$$I_{pn} = \frac{qD_n n_i^2 A_D}{L_n N_A}\left(e^{\frac{qV_{pn}}{kT}} - 1\right) \qquad I_{pn} \approx \frac{qD_n n_i^2 A_D^2 z_D}{L_n(N_A A_D z_D)} e^{\frac{qV_{pn}}{kT}}$$

The quantity in brackets: $y = N_A A_D z_D$ represents the varying doping on the low-doped side of the depleted diode. Here is where the random variation occurs. For this quantity the Poisson relation: $\mu_y = \sigma_y^2 = \lambda_y = N_A A_D z_D$ holds. Applying Eq. 5.11 gives for the variation in V_{pn} for a single diode:

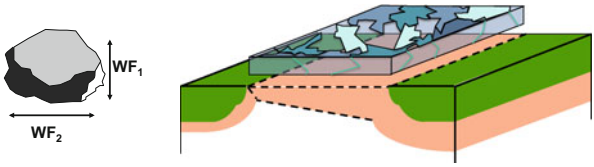


Fig. 5.33 In advanced technologies the gate is constructed from TiN or TaN grains. These metal-gates show work function variation [97]

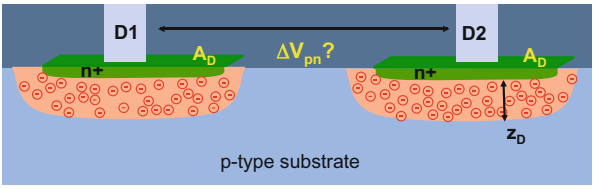


Fig. 5.34 Two diodes are fed with identical currents, how much do their voltages mismatch?

$$\sigma^2_{Vpn} = \left(\frac{dV_{pn}}{d\lambda}\right)^2 \sigma^2 = \frac{(kT/q)^2}{N_A A_D z_D}$$

and the difference between two diodes:

$$\sigma_{\Delta Vpn} = \frac{kT\sqrt{2}}{q\sqrt{N_A A_D z_D}}$$

Given that $N_A A_D z_D = 3000$ and $kT/q = 26\,\text{mV}$ the voltage difference between two diodes with identical currents is characterized by $\mu_{\Delta Vpn} = 0$ and $\sigma_{\Delta Vpn} = 0.67\,\text{mV}$. This simple derivation ignores some aspects: z_D is also dependent on the variation in doping. Also the random distribution in depth of the dopants is not considered. Note that a similar area dependence occurs as for the threshold voltage of an MOS transistor. The dopant level also plays an important role. A lower dopant level means that the diode voltage is created by having more distance between the dopants in the p- and n-region, as is found from inspection of the Poisson equation. The absence or presence of every ion has now a bigger impact on the diode voltage.

5.3.5 *Current Mismatch in Strong and Weak Inversion*

The matching properties of the current factor are derived by examining the mutually independent components W, L, μ, and C_{ox}, with the help of Eq. 5.11:

$$\frac{\sigma^2_{\Delta\beta}}{\beta^2} = \frac{\sigma^2_{\Delta W}}{W^2} + \frac{\sigma^2_{\Delta L}}{L^2} + \frac{\sigma^2_{\Delta C_{ox}}}{C^2_{ox}} + \frac{\sigma^2_{\Delta\mu n}}{\mu^2_n} \tag{5.18}$$

The mismatch generating processes for the gate oxide and the mobility are treated in accordance with Eq. 5.8. The variations in W and L originate from line-edge roughness. The analysis of edge-roughness is a one-dimensional variant of the analysis in the previous section and leads to $\sigma^2(L) \propto 1/W$ and $\sigma^2(W) \propto 1/L$.

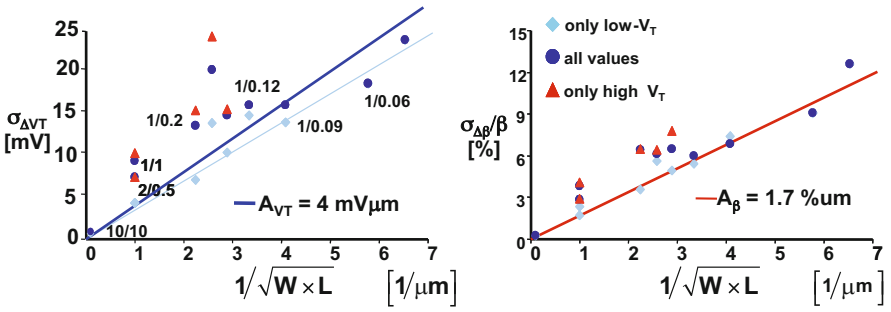


Fig. 5.35 The threshold mismatch and the relative current factor mismatch for NMOS transistors in 65-nm technology. Measurements by N. Wils/H. Tuinhout

$$\frac{\sigma^2_{\Delta\beta}}{\beta^2} = \frac{A^2_W}{W^2L} + \frac{A^2_L}{WL^2} + \frac{A^2_\mu}{WL} + \frac{A^2_{Cox}}{WL} \approx \frac{A^2_\beta}{WL} \tag{5.19}$$

where A_W, A_L, A_μ, and A_{Cox} are process-related constants.

A significant contribution of gate-oxide thickness variation would lead to a negative correlation between the threshold voltage mismatch and the current factor mismatch. Such a correlation is generally not observed. If W and L are large enough the respective contributions will also disappear. At gate-lengths below 65-nm, simulations [92] indicate some role for edge roughness. This role is in measurements hard to identify in the presence of large threshold mismatch. In [91] it was assumed that the matching of the current factor is determined by local variations of the mobility. Many experiments show that mobility affecting measures (e.g., mechanical pressure on the device) indeed lead to a strong increase in current factor mismatch. The relative mismatch in the current factor can be approximated by the inverse-area description as seen in the last part of Eq. 5.19. Some experimental data is shown in Fig. 5.35.

In contrast to the threshold random fluctuation, the absolute current factor variation is a function of temperature. However, the relative current factor mismatch as formulated in Eq. 5.19 is according to [95] much less sensitive to temperature.

Considering only the threshold and current factor variations,[8] the variance of the difference in drain currents ΔI_D between two equally sized MOS devices can be calculated. Using the generalized statistical method described by Eq. 5.11:

$$\sigma^2_{\Delta ID} = \left(\frac{dI_D}{dV_T}\right)^2 \sigma^2_{\Delta VT} + \left(\frac{dI_D}{d\beta}\right)^2 \sigma^2_{\Delta\beta} \tag{5.20}$$

[8]The contribution of mobility reduction factor θ and source and drain series resistance are next in line.

For strong inversion this equation is combined with the simple square-law current model for the current:

$$I_D = \frac{\beta}{2}(V_{GS} - V_T)^2 = \frac{W\beta_\square}{2L}(V_{GS} - V_T)^2 \tag{5.21}$$

$$\sigma^2_{\Delta I_D} = \beta^2 (V_{GS} - V_T)^2 \sigma^2_{\Delta VT} + \left(\frac{\sigma_{\Delta\beta}}{\beta}\right)^2 I_D^2 \tag{5.22}$$

$$\left(\frac{\sigma_{\Delta I_D}}{I_D}\right)^2 = \left(\frac{2\sigma_{\Delta VT}}{V_{GS} - V_T}\right)^2 + \left(\frac{\sigma_{\Delta\beta}}{\beta}\right)^2 \tag{5.23}$$

$$= \frac{2\beta_\square A^2_{VT}}{I_D L^2} + \frac{A^2_\beta}{WL} \tag{5.24}$$

At higher gate voltages the denominator term $V_{GS} - V_T$ will reduce the contribution of the threshold mismatch in the total current. Current mismatch goes mainly down with large L. The limited effect of increasing W can be understood by considering that indeed $\sigma_{\Delta VT}$ reduces, but unfortunately the transconductance goes up with the same factor. It is advisable to use high gate-drive voltages and corresponding low W/L ratios to minimize current mismatch. This choice reduces the transconductance and thereby also reduces noise, $1/f$ noise, and other spurious signals such as cross-talk on the bias lines.

In a circuit with constant current levels, the denominator term $V_{GS} - V_T$ of Eq. 5.25 will increase with rising temperatures. Under the assumption that the threshold mismatch and the relative beta mismatch remain constant with rising temperature [95], this will result in less the current mismatch.

In current-steering digital-to-analog converters, Sect. 7.3.2, the variance of single current sources is often used. This variance is half of the variance of the difference between the currents in two identically designed transistors as in Eq. 5.24:

$$\left(\frac{\sigma_{I_D}}{I_D}\right)^2 = \frac{1}{2}\left(\frac{\sigma_{\Delta I_D}}{I_D}\right)^2 = 2\left(\frac{\sigma_{\Delta VT}}{V_{GS} - V_T}\right)^2 + \frac{1}{2}\left(\frac{\sigma_{\Delta\beta}}{\beta}\right)^2 = \frac{2A^2_{VT}}{WL(V_{GS} - V_T)^2} + \frac{A^2_\beta}{2WL} \tag{5.25}$$

The input referred mismatch σ_{Vin} is often used in opamp and comparator designs. All disturbing factors, in this case the threshold and β mismatch, are summed in an equivalent input error:

$$\sigma^2_{Vin} = \frac{\sigma^2_{\Delta I_D}}{g^2_m} = \sigma^2_{\Delta VT} + \left(\frac{\sigma_{\Delta\beta}}{\beta}\right)^2 \left(\frac{V_{GS} - V_T}{2}\right)^2 \tag{5.26}$$

Reducing the **relative** current mismatch in Eq. 5.25 requires to increase the drive voltage $(V_{GS} - V_T)$, while lowering the **absolute** current mismatch and the input referred mismatch in Eq. 5.26 necessitates a low drive voltage. The circuit designer should carefully examine what mode optimally suits a circuit.

Equation 5.25 suggests an infinite current mismatch if the gate voltage approximates the threshold voltage. However, in that mode of operation the weak inversion model is applicable and levels off the maximum current mismatch. In weak inversion the current is modeled as an exponential function. Due to the low current level, the current factor mismatch is of less importance. Applying the Taylor approximation [9] gives the mean current and the variance of the resulting log-normal distribution for the current:

$$I_D = I_0 e^{\frac{q(V_{GS}-V_T)}{mkT}} \tag{5.27}$$

$$\mu_{ID} = I_0 e^{\frac{q(V_{GS}-V_T)}{mkT}} \left(1 + \frac{1}{2}\left(\frac{q\sigma_{\Delta VT}}{mkT}\right)^2\right) \tag{5.28}$$

$$\left(\frac{\sigma_{\Delta I_D}}{I_D}\right)^2 = \left(\frac{q\sigma_{\Delta VT}}{mkT}\right)^2 \tag{5.29}$$

Note that the mean value is larger than the nominal current without mismatch due to the asymmetrical behavior of the exponential function.

Figure 5.36 shows an example of the current mismatch relative to the drain current. At high gate–source voltages the current factor mismatch in the strong inversion equation 5.25 dominates. At lower gate–source voltages the threshold-related term in this equation gains importance. Except for extremely low current densities where the depletion layer width is shrinking, the observation can be made that the same value for $\sigma_{\Delta VT}$ applies for both the strong and the weak inversion regime. This example shows that the operating regime, where the mismatch parameters of the transistor are extracted, has a marginal effect on the accuracy[9] of the prediction in other regimes, as confirmed in, e.g., [88, 89, 98].

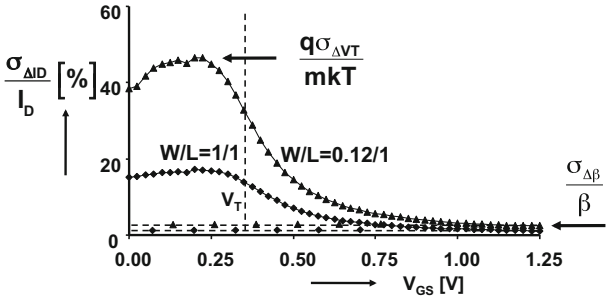


Fig. 5.36 The relative current mismatch for two 65-nm technology transistor geometries swept over the full voltage range. Measurements by N. Wils/H. Tuinhout

[9]Accuracy means that the standard deviation of a circuit parameter is within 5–10 % of the prediction, see Sect. 5.3.8.

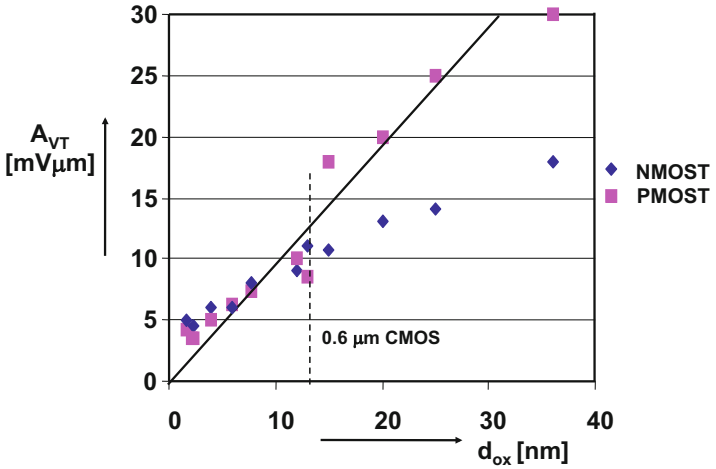


Fig. 5.37 Development of the threshold mismatch factor A_{VT} for NMOS and PMOS transistors as a function of the nominal oxide thickness of the process. The processes span 45-nm up to 1.6 μm CMOS

In a 65-nm process the mismatch factor $A_{VT} = 3.5\,\text{mV}\mu\text{m}$. So $\sigma_{VT} = 10\,\text{mV}$ for a 0.12/1 device. Equation 5.28 predicts a relative current mismatch of 40 %, which is confirmed by the measurement in Fig. 5.36. For a Gaussian distribution this would imply a reverse drain current below -2.5σ. At these levels of mismatch the assumption that a small Gaussian distributed threshold mismatch voltage will turn into a Gaussian approximation of the mismatch current is not valid anymore. The probability density function of the current mismatch needs here a log-normal distribution.

5.3.6 *Mismatch for Various Processes*

In Fig. 5.37 the threshold mismatch coefficient A_{VT} is plotted as a function of the nominal oxide thickness. As predicted by Eq. 5.16 [87, 91]:

$$\sigma_{\Delta VT} = \frac{A_{VT}}{\sqrt{WL}} = \frac{q d_{ox}\sqrt{2N_x z_d}}{\epsilon_{ox}\sqrt{WL}} \rightarrow A_{VT} \propto d_{ox}\sqrt[4]{N_x} \tag{5.30}$$

The mismatch coefficient is less for thinner gate-oxide thickness. Of course many more changes in the device architecture took place, still the oxide thickness seems to be the dominant parameter. The large PMOS transistor coefficients for >0.6 μm CMOS generations are caused by the compensating implants: the N- and PMOS transistor threshold adjust and n-well implants. The quantity $N_x = (N_a + N_d)$ is relevant for matching, while the net value $(N_a - N_d)$ determines the threshold in the

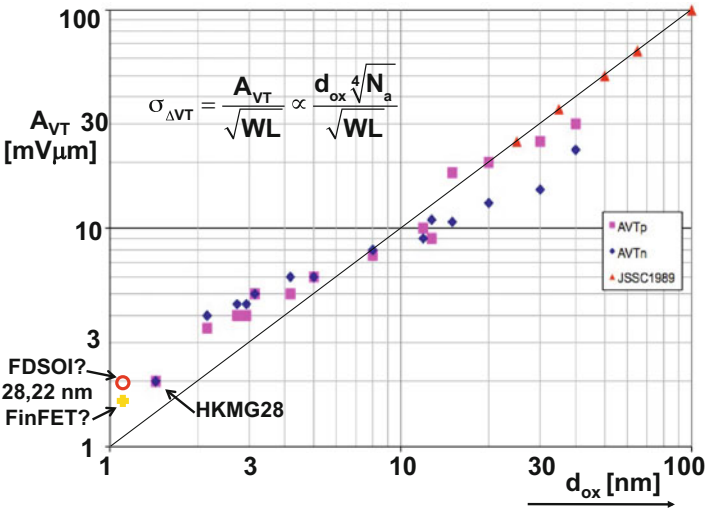


Fig. 5.38 The threshold mismatch factor A_{VT} for NMOS and PMOS transistors as a function of the nominal oxide thickness of the process on log–log scale. The processes span 28-nm up to 5 μm CMOS [74, 91]

PMOS transistor. Beyond the 0.6 μm node a twin well construction with a dedicated well implant for the PMOS transistor is used that avoids compensating charges.

In Fig. 5.38 the diagonal line indicates an A_{VT} factor increase of 1 mVμm for every nm of gate insulator thickness. This line is a first order estimate of what a well-engineered process should bring. Over the same process range the current mismatch factor A_β varies between 1.2 and 2 %μm.

The first indications of high-k factor metal-gate and Finfet processes are favorable. With $A_{VT} \approx 2$ mVμm the trend is continued. Nevertheless some concern is present on the increased threshold gradients in these processes as well as the variability of the source–drain diffusion resistances.

Figure 5.39 compares the MOS mismatch to the development of power supply voltage. A transistor with a size of $1.5L^2_{min}$ was chosen. During the process development from the 2.5 μm process to the 0.35 μm process both the mismatch and minimum gate-length have reduced. The power supply remained fixed at 5 V for the micron range process generations. Circuits that relied on analog CMOS performance such as analog-to-digital converters could improve their performance in these process generations by not fully following the line-width reduction.

At the 0.35 μm CMOS node the maximum electrical fields in intrinsic transistors were reached for both the vertical gate-oxide field and the lateral field controlling the charge transport. For this reason and in order to reduce power consumption, the power supply voltage was lowered all with the dimensions. On the other hand, the need to tailor the internal fields in the transistor has led to less uniform and higher implantation channel dope. As can be expected from the theoretical background, the

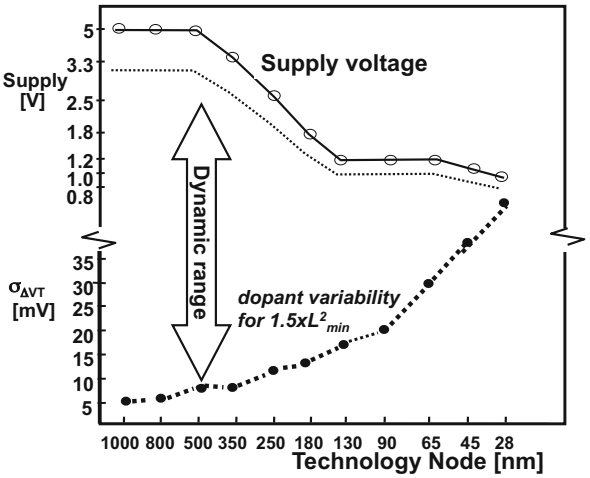


Fig. 5.39 Development of power supply voltage, signal swing (*dotted line*), and the measured NMOS threshold matching of a transistor 1.5 times the minimum size through various process generations

slower scaling of the gate-oxide thickness made that the threshold matching factor A_{VT} stopped decreasing. This becomes especially pronounced in 65–32 nm technologies, where pocket implants create an additional mismatch source. Shrinking the area of analog blocks in sub-micron processes is clearly an important economical issue, but in combination with a rising mismatch coefficient this will lead to lower performance. The reduction in the signal-to-matching coefficient ratio in sub-micron CMOS will necessitate changes in the system, design, or technology. In order to maintain high-quality signal processing, some enhancements to the standard processes are needed, such as the use of high-voltage devices or precision resistors and capacitors.

The effect of mismatch in advanced processes is further illustrated in Fig. 5.40 where the mismatch of a minimum size transistor in each process node is shown. It is obvious that below the 0.18 μm node a serious increase of mismatch takes place. Given the simultaneous decrease of the power supply voltage, transistor mismatch becomes a major bottle-neck in sensitive designs.

Example 5.6. In Fig. 5.41 four transistors circuits are given. Calculate the mismatch if $A_{VT} = 10\,\text{mV}\mu\text{m}$.

Solution. The upper left circuit contains two transistors each 20/5 μm. As A_{VT} has been defined for the difference between two transistors, the standard deviation can directly be calculated: $\sigma_{Vin-Vref} = A_{VT}/\sqrt{WL} = 10/\sqrt{5 \times 20} = 1\,\text{mV}$.

The upper-right circuit in Fig. 5.41 contains twice the number of components, based on the observation that M parallel circuits will show a $1/\sqrt{M}$ reduction in

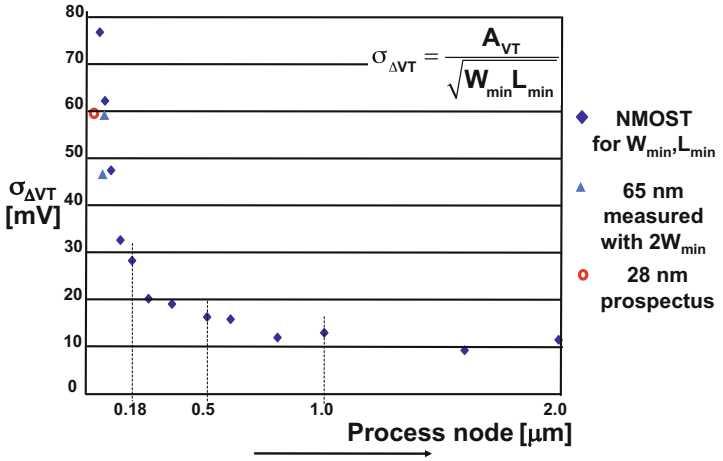


Fig. 5.40 The threshold mismatch of a pair of minimum size transistors over process generation following the process-design kit. The *diamonds* are taken from the design manuals for minimum size width and length of an NMOS transistor. The *triangles* represent actual measurements of $2W_{min}/L_{min}$ side transistors

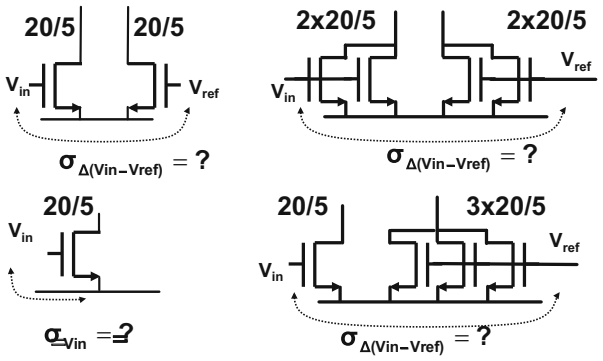


Fig. 5.41 Four mismatch situations

standard deviation, the solution is found as $\sigma_{Vin-Vref} = 0.7\,\text{mV}$. The same result is obtained by summing up the total area, and use the standard formula.

The lower left transistor is not part of a pair. Nevertheless a standard deviation of its threshold voltage (or any derived circuit parameter) can be specified by considering that two of these transistor in series via their sources form the first circuit. As the standard deviations of two transistors are mutually uncorrelated, each must amount 0.7 mV in order to get 1 mV for a series connection.

Table 5.6 An overview of matching models and value ranges.

MOS transistors	$\sigma_{\Delta VT} = \frac{A_{VT}}{\sqrt{WL}}$	$A_{VT} = 1\,\text{mV}\mu\text{m/nm}$ [74, 91]
MOS transistors	$\frac{\sigma_{\Delta\beta}}{\beta} = \frac{A_\beta}{\sqrt{WL}}$	$A_\beta = 1\text{–}2\,\%\mu\text{m}$ [74, 91]
Bipolar transistors (BiCMOS)	$\sigma_{\Delta Vbe} = \frac{A_{Vbe}}{\sqrt{WL}}$	$A_{Vbe} = 0.3\,\text{mV}\mu\text{m}$ [90]
Bipolar SiGe		$A_{Vbe} = 1\,\text{mV}\mu\text{m}$ [99]
Diffused/poly resistors	$\frac{\sigma_{\Delta R}}{R} = \frac{A_R}{\sqrt{WL}}$	$A_R = 0.5/5\,\%\mu\text{m}$
Plate, fringe capacitors	$\frac{\sigma_{\Delta C}}{C} = \frac{A_C}{\sqrt{\text{C in fF}}}$	$A_C = 0.3 - 0.5\,\%\sqrt{\text{fF}}$ [78]
Small capacitors <2 fF		$A_C = 0.7\%\sqrt{\text{fF}}$ [82]

The last circuit uses the previous result. Now the standard deviation of a single transistor must be added to the parallel connection of three. With the previous result in mind, this is a series connection of a single device and three devices in parallel, so: $\sigma_{Vin-Vref} = \sqrt{0.7^2 + 0.7^2/3} = 0.81$ mV.

5.3.7 Application to Other Components

In Table 5.6 matching parameters of various components are listed. In the previous paragraphs the mismatch parameters for the MOS transistor have been extensively discussed.

The behavior of the bipolar transistor is dominated by the number of dopants in the base that are not depleted. The fluctuation of this number, comparable to the fluctuation of the charge in the depletion layer of the MOS transistor, causes the base-emitter voltages between two bipolar devices to mismatch. Therefore a variance can be defined for ΔV_{be}. In [90] various experiments have confirmed the validity of this mismatch model.

Resistors for high-precision analog design are formed by polysilicon or diffused n- or p- doped areas. In advanced processes these layers are covered with a silicide layer to lower their impedance to the 2–5 $\Omega/\square$ level. A special mask is applied to prevent the deposition of silicide in order to obtain sheet resistances of 50–500 $\Omega/\square$.

Polysilicon resistors are enclosed by silicon dioxide acting as a thermal isolator. Dissipated heat may destroy such a resistor, a lower amounts of dissipated heat will affect the grain boundary structure and lead to a lasting resistance value shift after cooling.

Resistors suffer from area related mismatch and from edge roughness. The general description for the relative mismatch is therefore:

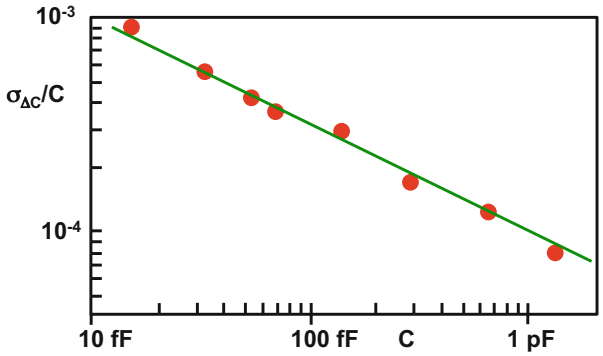


Fig. 5.42 Capacitor mismatch from [78], $A_{\Delta C} = 0.3\%/\sqrt{fF}$

$$\frac{\sigma_{\Delta R}^2}{R^2} = \frac{A_W^2}{W^2 L} + \frac{A_L^2}{W L^2} + \frac{A_\mu^2}{WL} \approx \frac{A_R^2}{WL} \tag{5.31}$$

The mobility variation mechanism includes impurity/doping scatter and in the case of polysilicon resistors also includes grain boundary disturbance. The last mechanism is important as the mismatch coefficient A_R increases to 5 %µm, while the diffused resistors allow to use 0.5 %µm. An additional factor in resistor design are the head-end connections [83, 84]. These connections introduce edge roughness, but more importantly also impose stress on the entire structure. A careful design or an additional margin is needed.

In [80] mismatch of capacitors was attributed to edge effects and to area non-uniformities. The first error would result in a line variation $\sigma_{\Delta C} \propto \sqrt{L}, \sqrt{W}$. The area non-uniformities comply to the mathematical conditions of the general model resulting in:

$$\frac{\sigma_{\Delta C}}{C} = \frac{A_{C,area}}{\sqrt{\text{area}}} = \frac{A_{C,area}\sqrt{\varepsilon_0 \varepsilon_r / d_{ox}}}{\sqrt{C}} = \frac{A_C}{\sqrt{C}} \tag{5.32}$$

At small dimensions the edge effects will dominate [82], but at reasonable sized capacitors >0.1 pF, the area effects become dominant [78] until the capacitor size becomes so large that wafer scale variations dominate (Fig. 5.42). The description of capacitor mismatch in Table 5.6 is different from the resistor model. For resistors the W/L ratio determines the resistor value, allowing to choose the device area independent of the resistor value. The capacitor value is always proportional to the area WL.

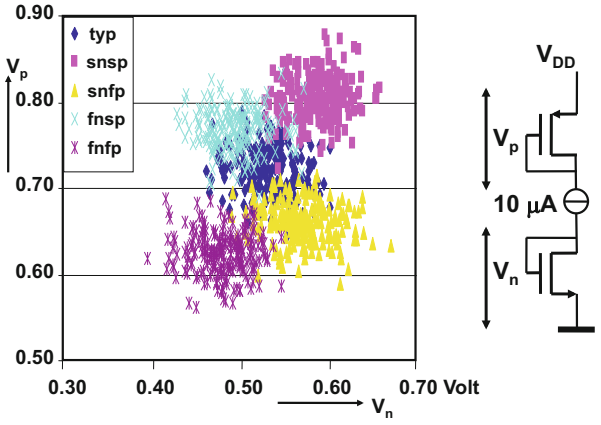


Fig. 5.43 Simulation of 200 0.2/0.1 P- and NMOS transistors in their 90-nm process corners. The notation "snfp" refers to slow NMOS and fast PMOS transistors. The variation due to mismatch is of equal importance as the process variation

5.3.8 Modeling Remarks

Figure 5.43 shows a simulation of the threshold voltage of 200 NMOS and PMOS 90-nm devices in their process corners. Although the corner structure is still visible, it is clear that for small transistors in advanced processes the mismatch is of the same magnitude as the process corner variation. The plot is somewhat misleading as it combines global process corner variation and local mismatch. A simple "root-mean-square" addition of these two variation sources ignores the fundamental difference between the two.

The simulation exemplifies that the random mismatch component for small devices is comparable to the global standard process-design kit parameter variations. Some designers even believe that simulating the nominal case with random parameter variation is sufficient given today's high-quality manufacturing capabilities. As a consequence the efforts to improve the quality of the random parameter simulations have increased. CAD vendors offer special simulation tools. Moreover an accurate model is required.

The present model describes random variations of devices by means of two statistical variables on the device-parameter level. This intuitive description allows easy communication between foundries, designers, and architects to balance power, area, and performance of mismatch-sensitive circuits. The standard deviations of Monte-Carlo simulated circuit parameters and from measured circuits agree within approximately 10 %, see, e.g., Fig. 6.11. This number is based on quantities of around hundred measured samples of a circuit fabricated in a stable process, without foundry liability margins. A significantly better accuracy requires quadratically more samples and, e.g., corrections for the effective channel width and length.

In [91] the mismatch contributions of the back-bias factor and the mobility reduction coefficient have been added. Other authors [86, 93] have proposed methods to get a closer prediction of mismatch in the various operating modes. One problem is that extra mismatch factors are not easy separable from others, requiring high-precision measurements [100]. For example, the mismatch in the mobility reduction coefficient θ and the source series resistance R_s are to a certain extent interchangeable: $\sigma(\theta) \Leftrightarrow \sigma(\beta R_s)$.

From a mathematical perspective, the threshold mismatch coefficient in an $I_D - V_{GS}$ plot corresponds to the variation in the back-extrapolated zero crossing of the drain current in the linear region, while the current factor mismatch coefficient is derived from the slope of the curve. So variations in the $I_D - V_{GS}$ plot are characterized by the zero order and first order mismatch components. Other potential mismatch causes are absorbed in threshold voltage and current factor mismatch coefficients. Identifying more mismatch effects requires a large statistical effort to realize sufficient significance.

Alternative model approaches concentrate on the physical parameters, e.g., mobility instead of current factor. Michael and Ismail [86] analyzes the individually extracted (BSIM) model parameters of test devices by means of a principal component analysis (PCA). This technique assigns the overall mismatch to the most relevant model parameter and then re-iterates to assign the remaining variation to other model parameters. Depending on the data points, this approach can lead to unexpected cross-correlations with little physical significance. In [101] the measured I-V plots are directly used as the input for a PCA analysis. A large numerical effort seems imperative. The work indicates that just a few principal components sufficiently describe the statistical behavior [101, Fig. 5]. This observation is in line with, e.g., [102].

5.4 Consequences for Design

A designer has several options to deal with global variations in his circuit. Next to rigorous simulation, the circuit itself can be designed in a way that the impact of global variations is minimized. These techniques include differential design and replica-biasing. Unfortunately these methods are ineffective for battling local variations. Here the designer has to rely on statistical predictions.

The threshold and current factor mismatch model that were defined in the previous sections describes the statistical properties of random mismatch in devices. In the design phase of a circuit these abstract notions of statistical variations on parameters must be translated into design decisions. Application to electronic design includes techniques as Monte-Carlo simulation, hand calculation, and various other techniques.

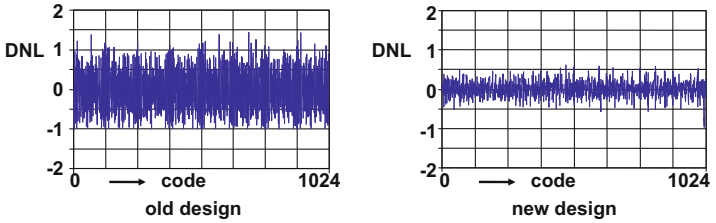


Fig. 5.44 These measured differential non-linearity plots are a quality parameter of this 10-bit analog-to-digital converter. *Left*: first prototype, *right*: second silicon after simulation-supported mismatch tuning of the circuit

5.4.1 Analog Design

Analog design relies heavily on mutual equality of identically designed devices. In Sect. 6.2.4 bandgap circuits are discussed. The caveat in the circuit is in the random offset of the operational amplifier input. This offset is also amplified and will create significant differences in the output voltage of the circuit. In Fig. 6.11 the measured output voltages of a few hundred devices are compared to the Monte-Carlo simulation of a similar number of samples. The MOS model in the simulation is equipped with a mismatch model as in Eq. 5.16.

Of course the following chapters will show many examples of the impact of mismatch on analog-to-digital converters and digital-to-analog converters. A first example is shown in Fig. 5.44. The differential non-linearity curve (DNL) is a measure for the error that an analog-to-digital converter makes at every code transition. If the DNL reaches the ± 1 limits, unacceptable deviations in the analog-to-digital transfer curve occur. In high-speed converters MOS threshold mismatch of the input pair of the comparators is the dominant contributor to differential non-linearity. It is imperative that this error must be made small. The measurement of the first prototype in Fig. 5.44 (left) shows significant deviations with excursions down to -1 (non-monotonicity). After analyzing the design with a Monte-Carlo simulation, the major contributors were located and correctly dimensioned, resulting in the measured curve on the right. The Monte-Carlo analysis is a standard tool for designing high-performance circuits in analog design.

The consequences of mismatch on the yield of analog-to-digital converters are discussed in Sect. 8.2.2. The shape of the curve in Fig. 8.32 is characteristic for yield plots. The transition from excellent yield numbers to commercially uninteresting levels is sharp. It shows importance of using the correct models and parameter values in precision designs.

Example 5.7. What improvement in DNL can you expect for a matching-sensitive ADC if the oxide thickness of a process reduces by a factor 2? And what is the improvement if also the power supply (and signal amplitude) decreases by a factor 2?

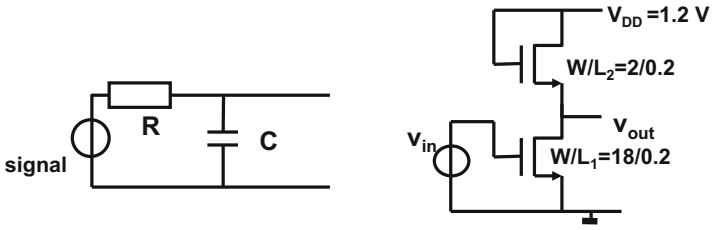


Fig. 5.45 Two circuits suffer from mismatch

Solution. Inspecting Eq. 5.16 shows that the $\sigma_{\Delta VT}$ is proportional to the oxide thickness. As $\sigma_{\Delta VT}$ is of direct impact on the DNL, a reduction of the oxide thickness by a factor 2 will improve the DNL with a factor 2. A reduction of both the oxide thickness and the signal amplitude by a factor 2 (the so-called constant-field scaling), does not change the DNL!

Example 5.8. In a first order R-C network, Fig. 5.45, the relative standard deviation of the resistor is $\sigma_R/R = 4\%$ and the relative standard deviation of the capacitor is $\sigma_C/C = 3\%$. What is the relative standard deviation on the bandwidth?

Solution. The bandwidth is linked to the values of the resistor and capacitor as: $BW(R, C) = 1/2\pi RC$.

Now Eq. 5.11 can be used to evaluate the bandwidth variance:

$$\sigma^2_{BW(R,C)} = \left(\frac{\partial BW(R,C)}{\partial R}\right)^2 \sigma^2_R + \left(\frac{\partial BW(R,C))}{\partial C}\right)^2 \sigma^2_C$$

$$= \left(\frac{-1}{2\pi R^2 C}\right)^2 \sigma^2_R + \left(\frac{-1}{2\pi RC^2}\right)^2 \sigma^2_C = BW^2(R,C)\left(\frac{\sigma^2_R}{R^2} + \frac{\sigma^2_C}{C^2}\right)$$

Substituting the values gives a relative bandwidth mismatch of 5%.

Example 5.9. Calculate the variance of the output voltage of the amplifier in Fig. 5.45. The input voltage is 0.5 V, and the threshold of both transistors is 0.4 V. $\beta_\square$ is unknown, $A_{VT} = 4\,\text{mV}\mu\text{m}$.

Solution. First the standard bias point calculations must be made. The transistors have the same current. So by equating both transistor currents:

$$I_{DS} = \frac{\beta}{2}(V_{GS} - V_T)^2 = \frac{W_1\beta_\square}{2L_1}(V_{GS_1} - V_T)^2 = \frac{W_2\beta_\square}{2L_2}(V_{GS_2} - V_T)^2$$

With the given parameters: $V_{GS2} - V_T = 0.3$ V, now $V_{out} = V_{DD} - V_{GS2} - V_{T2} = 0.5$ Volt. The gain of this stage is easily found from the above equation and amounts $A_v = -3$.

The output voltage fluctuates due to the threshold variations of both transistors. The fluctuations are

$$\sigma_{VT1} = \frac{A_{VT}}{\sqrt{2W_1L_1}} = 1.5\,\text{mV} \qquad \sigma_{VT2} = \frac{A_{VT}}{\sqrt{2W_2L_2}} = 4.5\,\text{mV}$$

As the single transistor variance is now relevant, the standard mismatch formula that describes the mismatch of a pair of transistors is modified by dividing by $\sqrt{2}$. Now the result is found by considering that the threshold mismatch of transistor 1 is amplified as if it is a small signal, while the threshold variations of the load transistor 2 directly affect the output voltage, so:

$$\sigma_{Vout} = \sqrt{A_v^2\sigma_{VT1}^2 + \sigma_{VT2}^2} = 6.3\,\text{mV}$$

The input related variance is: $\sigma_{Vin} = \sigma_{Vout}/A_v = 2.1$ mV. The same result is obtained by applying Eq. 5.11 on the current equation.

5.4.2 *Digital Design*

Also digital designers experience that for small devices the random component can exceed the process corner variation. An example is shown in Fig. 5.46 and Table 5.7 [103]. A pulse is applied to two sets of inverters and ideally it is expected that the outputs will change state simultaneously. Due to mismatch between the transistors of both rows of inverters, a random skew will exist in the difference of arrival time

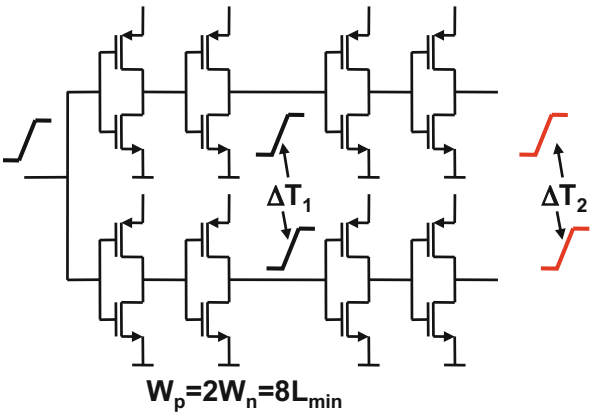


Fig. 5.46 An input pulse is applied to two chains of inverters. Due to mismatches in the transistors there is a time difference at the two outputs. This example mimics a critical timing path

Table 5.7 The simulated standard deviation of the difference in arrival time of the two pulses in the inverter chain of Fig. 5.46

Process node	0.25 µm	0.18 µm	0.13 µm	90 nm	65 nm
Clock period	10 ns	5 ns	2 ns	1 ns	0.5 ns
$\sigma_{\Delta T2}$ C_{load}=50 fF	16 ps	21 ps	38 ps	68 ps	88 ps
$\sigma_{\Delta T2}$ C_{load}=50..15 fF	16 ps	16 ps	22 ps	33 ps	32 ps

between various samples of this circuit. In Table 5.7 the standard deviation of this random skew is compared to the clock period in five technologies. From an effect in the 0.1 % range in 0.25 µm CMOS, the random skew[10] can take up to 10 % of the clock period in a 65 nm process.

In digital timing chains, an increasing chain length will linearly increase the deterministic skew, while the random component will go up via a square-root relation. The relative impact of random variations reduces. It should also be noted that the "root-mean-square" addition mechanism favors large contributors. The random timing skew in a long chain can be dominated by just one small-sized inverter. A similar analysis holds for ring oscillators in, e.g., PLLs.

Also in memory structures statistical effects play a major role. On the level of an SRAM cell, these effects can be evaluated as in any other analog circuit. The additional problem in memory structures is the millions of memory cells. This requires simulating statistical distributions up to 7σ. This is not practical in Monte-Carlo simulations. Special statistical acceleration mechanisms (importance sampling, Latin Hypercube sampling) allow to sample tails of statistical distributions [104]. Memory designs are affected by mismatch in several ways. Threshold variations influence the margins for the read and write operations. Moreover low-threshold devices create (exponentially) large leakage currents. The choice for the size of the transistors in an SRAM cell and in the sense amplifier critically depends on an accurate prediction of the mismatch coefficients.

Example 5.10. Three circuits in Fig. 5.47 experience threshold-voltage mismatch. Discuss how to reduce the effect of threshold mismatch for each case. Consider to increase or decrease the gate width or gate-length.

Solution. The left circuit in Fig. 5.47 shows a typical input stage for an opamp or comparator. Crucial is the input referred mismatch, which is caused by the threshold standard deviation. As this standard deviation is equal to $A_{VT}/\sqrt{WL}$ the area must be increased to reduce the mismatch. In most cases a large input transconductance is desired and the preferred option is to increase the gate width.

In the middle circuit the standard deviation of the output current must be decreased. Now it is important to realize that:

[10] Random skew is a random but time-invariant effect, while jitter is a random time-variant mechanism.

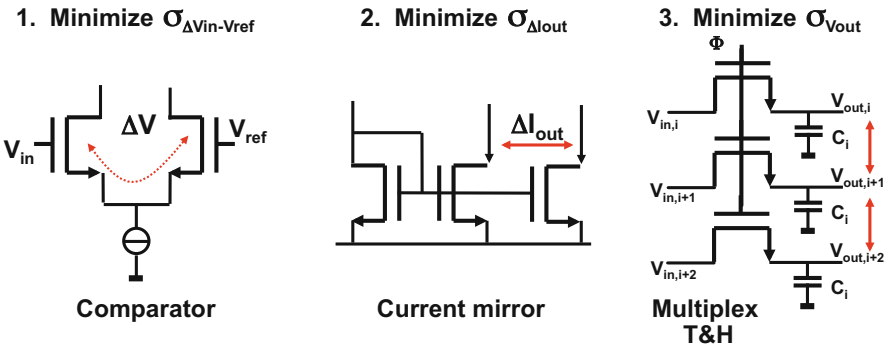


Fig. 5.47 Three circuits with mismatch

$$\sigma_{\Delta I} = g_m \sigma_{\Delta VT} = \sqrt{\frac{2W\beta_{\square} I}{L}} \frac{A_{VT}}{\sqrt{WL}} = \frac{A_{VT}}{L} \sqrt{2\beta_{\square} I}$$

Re-arranging terms makes clear that in first order increasing the gate width does not affect the result as the increase reduces the threshold mismatch but amplifies the remainder. So the solution is to increase the gate-length: the mismatch is reduced and the transconductance reduced.

In the last circuit three parallel track-and-hold circuits are implemented. It is crucial that all three track-and-hold circuits show the same output voltage for equal input voltage. The pedestal step due to the channel charge injection on the hold capacitor must be small and should not fluctuate per section. The standard deviation of the channel charge is $\sigma_{Qchan} = C_{gate}\sigma_{\Delta VT} = C_{ox}WL \times A_{VT}/\sqrt{WL}$. Evaluating the terms shows that σ_{Qchan} is minimum for the lowest gate area WL. So here the gate-length is certainly minimized to reduce the charge uncertainty and obtain the highest speed. The gate width is chosen at a value where the distortion is acceptable.

5.4.3 Drift

In literature the term "drift" refers to various phenomena in electronics circuits. Drift is a long-term parameter change that can be caused by aging effects, mechanical influences, and temperature. The most frequent appearance of drift in analog-to-digital conversion is a temperature-dependent shift of input offset. The magnitude of this drift is in the order of several μV/°C measured at differential inputs of circuits.

In a completely symmetrical and differential circuit with identical components there is no reason for drift. However, mismatch can cause unbalance that translates in drift. In a simple differential input pair fed with a constant current, inequalities will exist due to random threshold mismatch and current factor mismatch. The threshold mismatch is a charge based phenomenon and will create a temperature

independent offset [95]. This work also shows that the ratio $\Delta\beta/\beta$ is only marginally dependent on temperature. A current factor error will inflict a current error. With a constant current source feeding a differential pair, this error gets translated to a gate-drive change via the transconductance. The g_m contains a mobility term and is temperature dependent. This means that a temperature-independent $\Delta\beta/\beta$ error via a temperature-dependent transconductance causes drift at the input of a differential pair.

Assume that a current factor difference $\Delta\beta/\beta$ exists in a differential pair due to mismatch. The current factor difference $\Delta\beta/\beta$ creates a proportional offset current, which is translated into an input-referred offset ΔV_{in} via the transconductance:

$$\Delta V_{in} = \frac{I_d}{g_m}\frac{\Delta\beta}{\beta} \tag{5.33}$$

If a constant bias current is provided the only temperature dependence is due to the transconductance. As $g_m = \sqrt{2\beta I}$ and $\beta \propto T^{-\theta}$ with $\theta = 2, .., 4$, an expression for the temperature-dependent drift at the input of a differential pair due to current factor mismatch is found:

$$\text{Drift}=\frac{d\Delta V_{in}}{dT} = \frac{\theta}{T}\frac{I_d}{g_m}\frac{\Delta\beta}{\beta} = \frac{\theta}{T}\frac{V_{GS}-V_T}{2}\frac{\Delta\beta}{\beta} \tag{5.34}$$

5.4.4 Limits of Power and Accuracy

Power, speed, and accuracy span the design space, e.g., [105].

One of the main questions in low power conversion design is the ultimate limit of power consumption. Mathematicians would claim that the mapping of analog values on a discrete amplitude scale should not change the signal and therefore be zero-power.

In physics, however, a lower limit can be derived from quantum-mechanical effects, concerning the minimum number of electrons required to represent one bit. Another limit is given by Dijkstra [61] for $\Sigma\Delta$ converters. His analysis assumes that thermal noise on the input impedance or transconductance is dominant. This approach results in a energy limit based on the product of SNR and thermal kT noise and is the basis for the "Schreier Figure-of-Merit," Sect. 4.6.1. These limits are, however, three to four decades away from practical realizations. This is partly due to the fact that much "overhead" power has to be incorporated in real designs: transconductances in MOS need large amounts of current, parasitics have to be overcome, safety margins for all kinds of unexpected variations have to be accounted for.

The idea that accuracy must be balanced against power can be easily understood by considering that the threshold voltage uncertainty in combination with the gate capacitance can be described as an energy term [106, 107]. In this section an

approximation for power consumption in parallel-circuit structures as in high-speed converters will be derived based on random variations in the circuit components. This may be the case in multiplexed circuits or in circuits that rely on matching, as in full-flash converters. The component variations between the various paths or comparators can be interpreted as if an unwanted signal is added to an ideal structure: fractions of sample rates, fixed pattern noise, etc. These component variations will decrease when the area of an MOS transistor is increased. On the other hand, the loading of the signal will also increase when gate areas increases:

$$\begin{array}{ll} \text{Capacitive load} & \text{Threshold variance} \\ C_{gate} = WLC_{ox} & \sigma_{\Delta VT} = \dfrac{A_{VT}}{\sqrt{WL}} \end{array} \tag{5.35}$$

The voltage uncertainty on the gate capacitance can be described as an energy term [106, 107]:

$$E_{\sigma VT} = C_{gate}\sigma^2_{\Delta VT} = C_{ox}A^2_{VT} \approx 3 \times 10^{-19} \quad \text{Joule} \tag{5.36}$$

which is independent of the transistor size and corresponds to about 70 kT at room temperature. A similar analysis holds for flash converters in Fig. 5.48. This energy can be seen as the energy required to toggle a latch pair of transistors in meta-stable condition into a desired position with a 1σ certainty. In circuits with parallel data paths the energy required to overcome component mismatch may hence dominate thermal kT noise by two orders of magnitude.

This factor can be plotted versus the process generation, in Fig. 5.48.

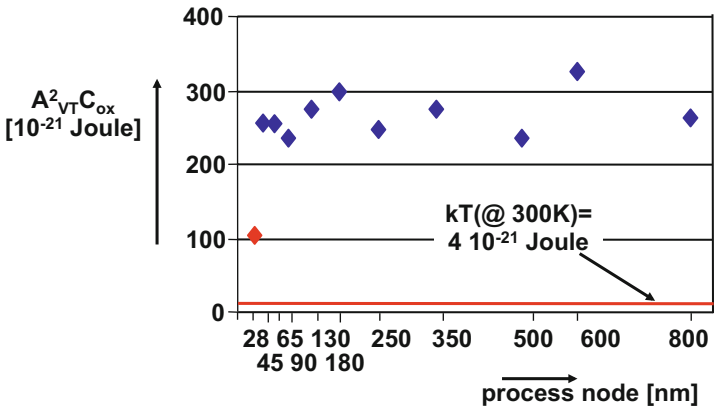


Fig. 5.48 The technology factor $A^2_{VT}C_{ox}$ versus process generation

It is clear that the energy associated with threshold mismatch hardly changes over time. Given the reducing power supply (from 3.3 V down to 1.1 V) this implies that the power efficiency of straight-forward full-flash architectures will not scale and perhaps even deteriorate.

Exercises

5.1. Explain what the consequences are of power supply reduction for analog circuits.

5.2. What is the reason in Example 5.3 on page 140 that the magnitude of the distortion decreases at higher signal frequencies.

5.3. The perimeter of a square unit capacitor contributes between 8 and 16 % depending on the batch processing (global variation). What is the accuracy of a 1:1.5 ratio, if this ratio is obtained by stretching one side of a capacitor.

5.4. An experimental current-steering digital-to-analog converter is based on a radix $e = 2.71\ldots$. Design four current sources with ratios: $1 : e : e^2 : e^3$. The mask grid allows only multiples of 0.1 μm and the structure cannot exceed an area of 100 μm^2.

5.5. In Example 5.4 on page 146 the currents in the individual transistors is 1 mA. The square resistance of the metal-1 is 0.1 Ω. Calculate the input referred mismatch.

5.6. Modify the lay-out of Fig. 5.27 in such a way that the sources are all towards the center and the drains are on the border of the structure. Is this an advantageous modification?

5.7. An SRAM latch with $W/L = 0.2/0.1$ transistor sizes for all transistors in 65-nm CMOS is designed. How much differential voltage is needed to reach a 6-σ probability that the latch toggles correctly from its meta-stable position.

5.8. An array of 64 NMOS current sources $W/L = 10/5$ in 90-nm CMOS is designed. A drive voltage of $V_{GS} - V_T = 200$ mV is applied. Calculate the standard deviation of the output current.

5.9. There is an area of 10 μm^2 available to generate from a perfect voltage source a current with minimum spread. Make a choice of components in 0.18 μm CMOS from Table 5.3. Give an estimate of the achievable accuracy.

5.10. In the previous exercise there is also an accurate clock signal available, enabling the use of capacitors defined in Table 5.2. Does the picture change?

5.11. Discuss the limits to the proposed length and width changes in Example 5.10 on page 171.

5.12. Discuss series resistors in power supply lines. Are they useful for decoupling?

Chapter 6
Reference Circuits

6.1 General Requirements

In[1] an analog-to-digital or digital-to-analog converter the reference value defines the analog signal range. The ratio between the analog signal and the reference value corresponds to the ratio of the digital signal to the maximum digital signal.

The reference quantity is mostly a voltage, current, charge, or time period. In the field of analog-to-digital conversion the reference value, and consequently the maximum input signal, of an N-bit converter is subdivided into $2^N A_{LSB}$, where A_{LSB} is the physical value corresponding to one least significant bit (LSB). During operation of an analog-to-digital converter or a digital-to-analog converter, this reference value is subdivided, copied, or multiplied, which causes various inaccuracies, see Chap. 5.

The reference quantity in the analog domain is therefore a key component for setting the quality and performance of analog-to-digital and digital-to-analog conversion. Any disturbance or error in the reference value will cause a deviation in the converted signal of the same relative magnitude.

A good reference quantity meets the following criteria:

- Reproducible in systems that are mass-manufactured: statistical variations must be minimized.
- Not sensitive to any (low-frequency or high-frequency) variation in the supplies. Next to the power supply also the effect of changes in surrounding potentials should not affect the reference value. For example, in case of an integrated circuit, voltage variations in the substrate must be considered.
- Immune to load variations: A reference has to have a sufficiently low source impedance. A reference with a relatively high output impedance will be unable

[1]The author thankfully acknowledges the corrections by Dr. Anne-Johan Annema, University Twente.

M. Pelgrom, *Analog-to-Digital Conversion*, DOI 10.1007/978-3-319-44971-5_6

to adequately react on variations in the current pattern, caused by the connected circuits. This load pulling can have multiple effects. Firstly the output signal of the reference circuit may shift due to non-linearities in the reference circuit, due to, e.g., rectification. Secondly pulling effects of a certain connected circuit directly propagate to all other connected circuits: spurious signals are then created and distributed. A common reason for malfunctioning or loss of performance is this sharing of a reference with inadequate output impedance over too many subcircuits.

- The reference should not introduce interference or noise into the surrounding circuitry. Most reference circuits use feedback circuitry. If no attention is paid to stability issues, or if a reference is loaded with an unexpected combination of capacitive and inductive elements, the reference circuit can superimpose on the reference quantity an oscillation.[2]
- Stable for temperature changes. Systems for industrial and consumer use are subject to temperature variations from $-20\,^{\circ}\mathrm{C}$ to $+120\,^{\circ}\mathrm{C}$ on the printed circuit board. In special cases these variations can be much higher (space, oil drilling) or much lower (pace-maker).
- Limited aging. This implies that the reference value should hardly change as a function of time. Aging of components is often related to high temperatures and over-voltage. Stressing a part by means of extreme temperatures and/or higher voltages can be used to extrapolate the expected life time and aging effects. Often the effects of the stress mechanisms show up a typical bathtub behavior for faulty behavior during the product life cycle. This means that on average a lot of defects occur both in the beginning and at the end of lifetime, while a rather stable behavior is obtained during the long period in between. In order to avoid this initial period with many defects, fresh products are often subjected to a short stress period to sort out the products that show these initial defects. This short stress period right after production is called the burn-in period.
- Low energy usage. The amount of energy that signal manipulations require depends on the complexity of the operation, the bandwidths, and accuracy. A reference is normally not listed in the high energy categories.

Note that reference sources are always equipped with two connections: the reference voltage is defined between two terminals. The current from the output of the reference circuit will flow from that output, via the load, back to the reference ground. Both terminals contribute equally to the performance of the reference circuit and in a design they should therefore be considered as being of equal importance.

Any reference value is based on some physical quantity. There are many physical quantities to choose from, see Table 6.1:

The determining factor for the choice is the degree in which a reference can fulfill the above requirements. A voltage division of a relatively quiet power supply may serve as a low-quality reference voltage. Zener and threshold related potentials can

[2]Don't say: "This will not happen to me."

Table 6.1 Various domains for choosing a reference

Domain	Example	Remarks
Voltage	Band-gap voltage, threshold voltage, Zener-voltage, power supply	Semiconductor physics related quantity
Current	Via a resistor derived from a voltage	Every current reference is a voltage reference in disguise
Charge	Electron, voltage on a capacitor	
Time	Crystal clock	For low frequency application

serve in simple applications. Accurate systems require special reference sources, and only a limited number of physical phenomena are sufficiently stable. Accurate time reference is derived from atomic resonance phenomena. Voltage references often use the potential difference between the conduction and the valence band in pure materials: the band-gap voltage V_{bg}. Industrial references based on a band-gap voltage can reach additional stability by using temperature compensation or by stabilizing their own temperature in a control loop. Regular calibration (e.g., yearly) further improves the performance.

In integrated circuits for consumer applications a simple band-gap reference presents a suitable solution for modest demands: an absolute accuracy of around 1 % and another 1 % variation during a temperature trajectory from 0 to 70 °C can readily be achieved. The band-gap reference idea was originally developed by Hilbiber [108] in 1964. Widlar [109], Kuijk [110], and Brokaw [111] refined the idea in the early 1970s.

6.2 Band-Gap Voltage Reference Circuits

6.2.1 Principle

The reference quantity in a band-gap voltage reference is the energy difference that an electron has to overcome for moving from the valence band in silicon to the conduction band [11]. In the following, a band-gap reference circuit in silicon is assumed. For silicon, the "band-gap voltage" varies from 1.17 V close to 0° K to 1.12 V at room temperature. This value of the band-gap voltage can be accessed via a circuit known as band-gap reference circuit that uses a number of physical properties of diodes in conjunction with circuit techniques. The diode current depends on energetic electrons that manage to jump over the band-gap barrier. The electrons have a thermal energy $kT/2$: at a higher temperature the electrons have proportionally more energy and (exponentially) more current will flow.

The current over a pn-junction is analyzed in detail in many solid-state physics textbooks, e.g., [11], and can be approximated[3] as:

$$I_{pn} = C_0 T^{-\eta} e^{\frac{q(V_{pn}-V_{bg})}{kT}} \tag{6.1}$$

Here, all temperature independent terms in Eq. 6.1 have been collected in the term C_0. Note that for normal forward operation of a diode $V_{pn} < V_{bg}$: hence the exponent is negative and $C_0 T^{-\eta} \gg I_{pn}$. The temperature behavior of $C_0 T^{-\eta}$ is mainly determined by the temperature dependence of the intrinsic carrier concentration n_i, and is additionally curved due to the temperature dependence of the mobility. Mathematically this temperature dependence is expressed via the term $T^{-\eta}$. Practical values of η range between 2 and 4. Now, to keep the diode current at a fixed value, an increase of temperature requires a decrease in V_{pn}. It can be derived that the pn-junction voltage V_{pn} shows a negative temperature coefficient of roughly $-2\,\mathrm{mV/°C}$.

The band-gap voltage at 0 K can be found by back-extrapolating the forward diode voltage from room temperature. This linear extrapolation results in a slightly higher value of 1.205 V because of some secondary effects in the band-gap voltage versus temperature curve.

Figure 6.1 shows the principle of a band-gap reference [108]: linear back-extrapolation of the diode voltage to 0 K is done by addition of a voltage $N_R \Delta V_{pn} \propto T$ to the diode voltage. This basic band-gap voltage reference circuit creates a relatively stable voltage equal to the linearly back-extrapolated band-gap of 1.205 V.

In a band-gap reference, the voltage $N_R \Delta V_{pn} \propto T$ is generated from the voltage difference between two pn-junction voltages with unequal current densities. The

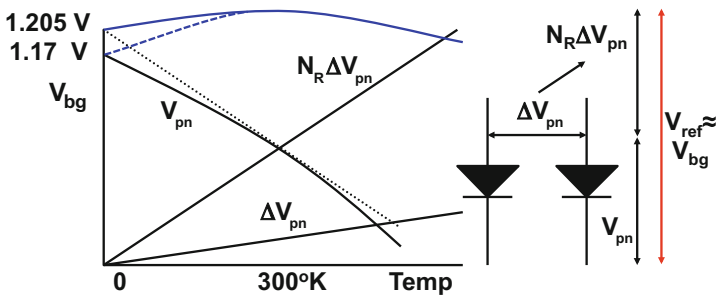


Fig. 6.1 The diode voltage V_{pn} decreases with increasing temperature at roughly 2 mV/°C. The differential voltage between two diode voltages ΔV_{pn} (e.g., created by different areas of the diodes) is proportional to the absolute temperature. By amplifying this last slope to a level where the differential voltage temperature coefficient is equal but opposite to the diode temperature coefficient, a temperature stable voltage can be obtained

[3]The correction factor to get $I_{pn}(V_{pn} = 0) = 0$ has been omitted.

current density is the current divided by the area of the pn-junction. The voltage difference ΔV_{pn} can be realized by feeding identical currents through a single pn-junction and through a parallel connection of N_D identical pn-junctions.[4] Using Eq. 6.1 for both diodes gives

$$\Delta V_{pn} = V_{pn,1} - V_{pn,ND} = \frac{kT}{q} \ln(N_D) \tag{6.2}$$

The voltage difference ΔV_{pn} shows a positive temperature coefficient, often referred to as "proportional to absolute temperature," PTAT. This property of two pn-junctions is not only the basis for band-gap reference circuits but it is also the basis for many temperature measurement schemes.

The term kT/q varies only with $86\,\mu\text{V}/°\text{C}$. Increasing this value with $\ln(N_D)$ to obtain 2 mV/°C is not feasible: a scaling factor of about 10^{10} between the two diodes would be required. As a resort, usually a manageable scaling factor is used, e.g., up to $N_D = 25$, for which $\ln(N_D) = 3.2$, and an opamp configuration will take care of additional multiplication.

Figure 6.2 shows a popular configuration: the circuit comprises an opamp, the single and N_D-multiple diode, moreover a small resistor R and two resistors with a value of N_R times the resistor R. The circuit is suited for bipolar, BiCMOS, and for CMOS implementation as in CMOS a parasitic diode is available between the diffusion of a PMOS transistor and the n-well diffusion, see Fig. 6.9.

The opamp will force its input terminals to a zero potential difference. Consequently there is an equal voltage drop over the two resistors $N_R R$, and the currents in the two branches are equal. As one branch has one pn-junction and the other branch N_D equally sized pn-junctions, a PTAT voltage difference as given by Eq. 6.2 exists between both pn-junctions. Because the opamp forces its inputs to be at the same

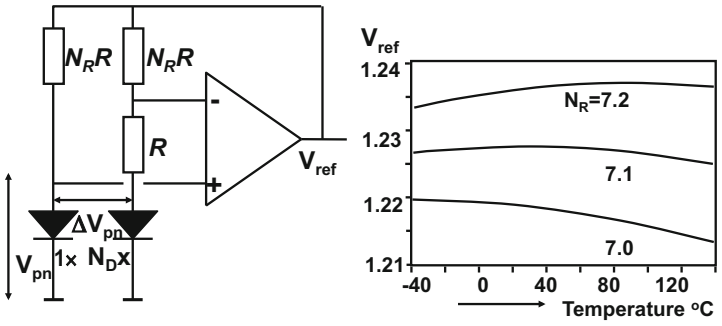


Fig. 6.2 The band-gap reference circuit [110]. The graph shows a simulation of this circuit with $N_D = 24$ and $N_R = 7.0, 7.1, 7.2$

[4] It is certainly possible to use two identical diodes and feeding two different currents. However, designing an accurate current ratio appears more difficult than designing an array of diodes.

potential, the voltage difference ΔV_{pn} is equal to the voltage drop over the resistor R. The resulting current through R flows also through both $N_R R$ resistors. Therefore the voltage over R is multiplied to $N_R \Delta V_{pn}$ over the top resistors; this voltage drop $N_R V_{pn}$ is PTAT. Assuming that the resistors are temperature independent,[5] this will also yield PTAT branch currents. The output voltage of the circuit V_{ref} is now:

$$V_{ref} = V_{pn} + N_R \Delta V_{pn} = V_{pn} + \frac{kT}{q} N_R \ln(N_D) \tag{6.3}$$

The design choice of $N_R \ln(N_D)$ allows to compensate the negative temperature coefficient of the pn-junction by the PTAT term. After substituting the current equation of a pn-junction, Eq. 6.1, in Eq. 6.3 the following equation for the reference voltage is obtained:

$$V_{ref}(T) = V_{pn} + N_R \Delta V_{pn} = V_{bg} + \frac{kT}{q} \ln(\frac{I_{pn}}{C_0 T^{-\eta}}) + \frac{kT}{q} N_R \ln(N_D) \tag{6.4}$$

The band-gap voltage V_{bg} in this equation is followed by a term that has a negative temperature coefficient because $I_{pn} \ll C_0 T^{-\eta}$ and is followed by a PTAT term created by the circuit. To obtain a reference voltage that is in first order independent of temperature, these two latter terms must cancel each other temperature dependence in first order.

After differentiation for the temperature T and setting the result to zero, a mathematical maximum value for the reference voltage is found at a temperature T_0:

$$\ln(\frac{I_{pn}}{C_0 T_0^{-\eta}}) - \eta + N_R \ln(N_D) = 0$$

Substituting this result in the original formula gives

$$V_{ref}(T) = V_{bg} + \frac{kT}{q} \eta (1 - \ln(\frac{T}{T_0})) \tag{6.5}$$

The maximum point in the reference voltage, at $T = T_0$, has as value

$$V_{ref}(T = T_0) = V_{bg} + \frac{kT_0}{q} \eta$$

Clearly, the choice of N_R determines both T_0 and determines the value of the maximum $V_{ref}(T = T_0)$, see Fig. 6.2. To estimate the curvature of the reference voltage, a series expansion of the logarithmic function in Eq. 6.5 can be made.

[5] Temperature dependence of the resistors or other components is canceled during the fine tuning of N_R in the simulation.

Around $T = T_0$ the argument of the logarithmic function is close to "1" and $\ln(1+\alpha) \approx \alpha$ for $|\alpha| \ll 1$. Substituting this in Eq. 6.5 yields

$$V_{ref}(T) = V_{bg} + \frac{kT_0}{q}\eta(1 - (\frac{T - T_0}{T_0})^2) \tag{6.6}$$

The output voltage of a band-gap reference voltage circuit has therefore a parabolic shape. The temperature T_0 at which the maximum occurs can be set by the resistor ratio N_R for a given diode ratio N_D. For example, if the diode ratio is $N_D = 8$, N_R will be around 11. The maximum output voltage is the band-gap voltage V_{bg} plus a few ($\eta \approx 2$–4) times the thermal voltage $kT/q \approx 26\,\text{mV}$. The typical output voltage is in the range 1.23–1.28 V for band-gap voltage references in silicon.

Figure 6.3 shows two band-gap reference circuit implementations employing an opamp. In the rightmost realization, both a PTAT current and a temperature stable reference voltage are available. Note that in that circuit, the output voltage is now out of the amplification loop, so all settling and slewing depends on the current in the last stage. Stability must be checked and guaranteed, although voltage references are DC circuits.

Example 6.1. A diode with a temperature coefficient of -2 mV/°C is supplied via a resistor R with a current of 1 µA at 27 °C. This current is linearly proportional to the absolute temperature (PTAT). At what value of R will the voltage over this combination show a temperature independent behavior?

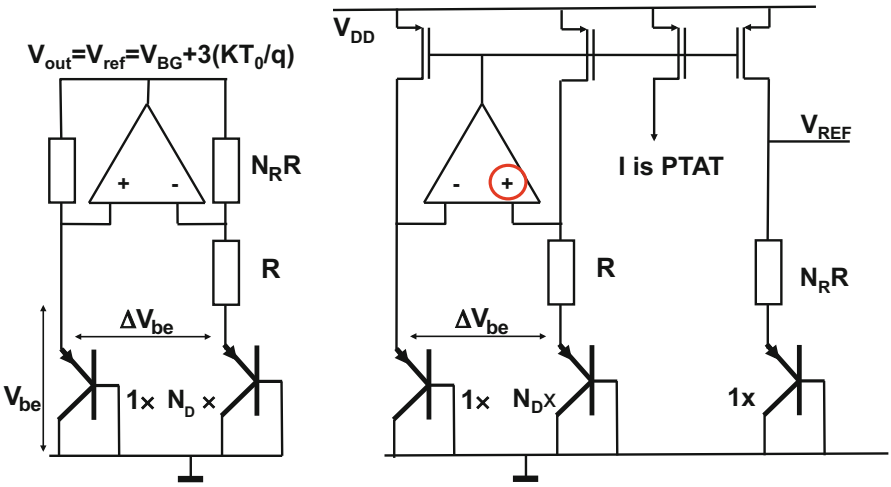


Fig. 6.3 *Left*: an often used way of drawing the standard reference circuit diagram has the opamp pointing upward. In the *right-hand side* schematic a simple modification allows to replace the resistors by current sources. The circuit delivers a current output proportional to absolute temperature (PTAT) and a reference voltage

Solution. The current can be written as $I_{PTAT} = 10^{-6} \times \text{Temperature}/300\,°\text{KA}$. Consequently the voltage over the resistor R will show a positive temperature coefficient that must balance the diode behavior:

$$\frac{I_{PTAT}R}{\text{Temp}} = 2\,\text{mV}/°\text{C}.$$

resulting in a resistor value of 600 kΩ with a voltage drop of 0.6 V at room temperature. Summed with the diode forward voltage, the combination will show an overall voltage drop of approximately 1.25 V.

6.2.2 *Artifacts of the Standard Circuit*

Band-gap voltage reference circuits are designed to have a stable point at and around the situation where their output voltage equals (approximately) the band-gap voltage. However, the vast majority of band-gap voltage reference circuits, like the circuit in Fig. 6.4, have a second stable point: at 0 V output voltage all branch current in the circuit is zero while the opamp input voltage is zero. Consequently these reference circuits need an additional start-up circuit to avoid this 0-V condition. A pull-up resistor on the reference voltage is often sufficient to force starting up. In Fig. 6.5 an NMOS transistor with its source connected to the reference circuit output, the drain to the power supply, and the gate to a voltage generated by two stacked threshold voltages will become active if the reference voltage is below one threshold voltage. More elaborate schemes switch off the start-up circuitry after sufficient output voltage is detected which usually decreases power consumption and increases accuracy.

Most of the problems of the standard circuit are related to the roughly tenfold voltage amplification in this circuit, required to make the PTAT voltage with temperature coefficient of about 2 mV/K. All noise and offset components related to the input of the operational amplifier (input related circuit effects) are multiplied

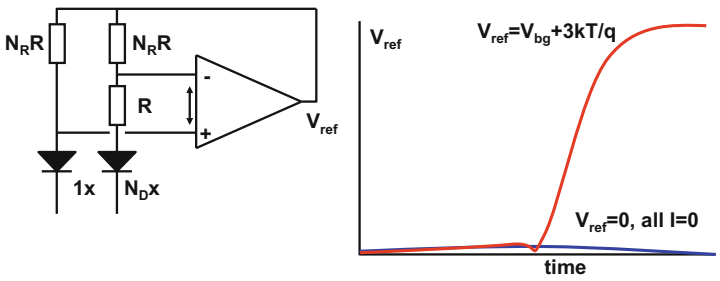


Fig. 6.4 A reference circuit has two stable output voltages: the desired reference voltage

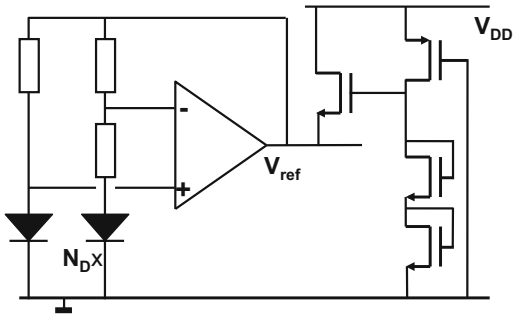


Fig. 6.5 A simple start-up circuit pulls the band-gap output away from 0-V. The transistor connected to the reference voltage will switch off as soon as the reference voltage rises above one threshold voltage

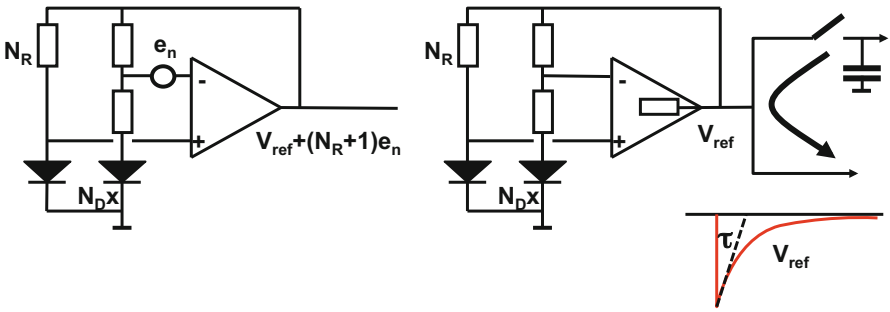


Fig. 6.6 A band-gap reference circuit amplifies unwanted input signals. If the output impedance is too high unwanted coupling will occur between connected circuits

with this factor, see Fig. 6.6 (left). Unwanted high-frequency components in the opamp output voltage can be suppressed by reducing the bandwidth of the opamp.

The disadvantage of a bandwidth reduction is a limited response of the circuit to any load variation, or suppression of any interference on the reference output terminal. Without sufficient bandwidth the reference circuit will not be able to respond these changes at the output. In other words: for those frequencies the output impedance is too high. A proper choice for the bandwidth requires considering the loading variations in the succeeding circuits. In some cases separate reference circuits for each separate block are preferred. Another solution is using a huge capacitor connected to the output of the band-gap circuit. Such an on-chip "super-capacitor" of tens of nanoFarads avoids problems with off-chip capacitors. External capacitors easily become ineffective as the required bonding wires show inductance of typically 1 nH per mm.

6.2.3 Bipolar Band-Gap Circuit

Only four bipolar transistors are needed to design a band-gap reference circuit [111], see Fig. 6.7. The base-emitter junctions of the two npn-transistors create the temperature sensitive pn-junctions. The pnp-mirror keeps the currents in both

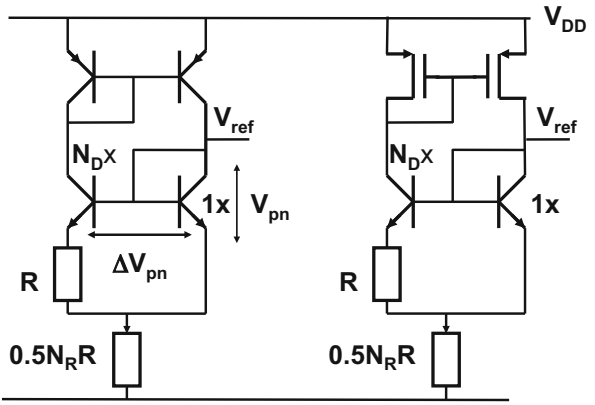


Fig. 6.7 A band-gap reference circuit with bipolar transistors and in a BiCMOS process

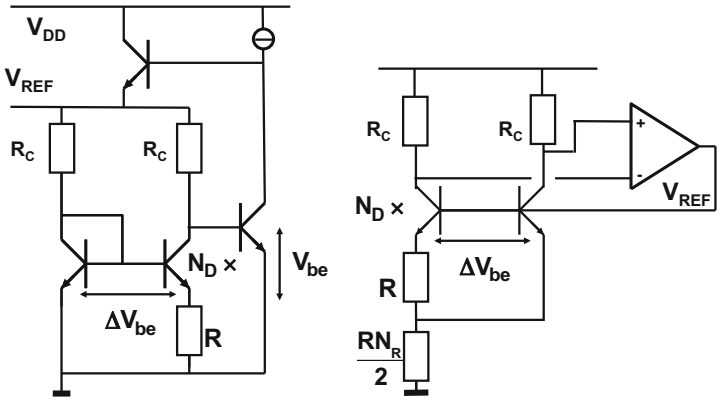


Fig. 6.8 *Left*: the Widlar reference circuit, *right*: the Brokaw reference circuit. These band-gap circuits use the bipolar transistors for amplification [109, 111]

branches equal. The npn-transistors in combination with the pnp-transistors form the operational amplifier. All currents in the circuit are PTAT. This circuit can be modified to act as a temperature sensor, for a detailed analysis, see, e.g., [112].

If all three terminals of the bipolar transistor can be used without restrictions as in Fig. 6.8, the amplification R_C/R of this class-A amplifier can be used to reduce the opamp noise and mismatch. Now the noise of resistors and bipolar transistors dominates.

In triple well CMOS, a parasitical npn is present with current gain of $h_{fe} = 6$–10. The collector construction as in Fig. 6.8 (right) still has advantages but introduces early voltage requirements, e.g., for PSRR, as well as requirements on the current gain factor h_{fe} matching.[6]

[6] see Stefan Marinca, ADI patents, e.g., US6.891.358.

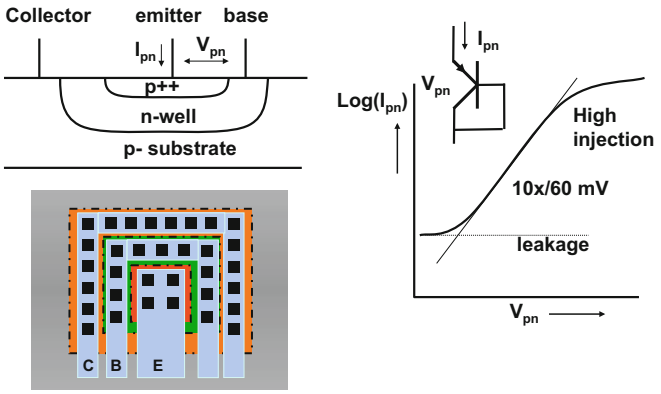


Fig. 6.9 The pn-diode or pnp transistor in a CMOS process and its I-V curve

6.2.4 CMOS Band-Gap Circuit

In a CMOS process the pn-junction is available as a parasitic structure created from a standard p-drain/source, an n-well, and the p-type substrate [113], see Fig. 6.9. In fact this structure is a parasitic pnp transistor where the substrate serves as a collector. Using this structure with forward current into the substrate also poses a danger: the hole current in the substrate may trigger latch-up. Latch-up in CMOS processes is possible if an intended or parasitic pnp structure feeds its collector current in an npn structure. A potential candidate for an npn structure is every n-MOS transistor, where the source, substrate, and drain form the npn structure. If this npn transistor starts conducting, the base current of the pnp will increase. Now an uncontrollable positive feedback process is started that can easily lead to excessive currents and damage to the circuits. Turning-on of the parasitic bipolar npn transistor requires on a sufficiently high base voltage (0.5–0.7 V). This base voltage can become high if there is a large resistor from the base to ground or the supply. It is therefore important to surround the npn and pnp-devices with sufficient substrate contacts to lower this resistance and to avoid the pnp collector current to build up a high base voltage for the npn devices.

Typical emitter sizes of the pnps for band-gap circuits in CMOS are $2 \times 2\,\mu m^2$ to $5 \times 5\,\mu m^2$. Due to the rather low-doped base (n-well) the current density cannot be large in this structure because then the diode experiences high-injection [11]. High-injection will cause deviation from the ideal I-V curve as shown in Fig. 6.9 that will typically result in reduced accuracy. Similarly, at very low current (density) levels, also deviations from the ideal I-V behavior occur, due to leakage and recombination. Typical band-gap voltage reference designs aim at a diode current range between $1\,nA/\mu m^2$ and $1\,\mu A/\mu m^2$.

The base width of this parasitic pnp transistor (0.3–0.5 μm) in CMOS is large compared to base widths in advanced bipolar RF technologies. High-speed bipolar devices normally use a base width of less than 0.1 μm and this relatively short

distance that the minority carriers must travel allows to operate them at high frequencies. For band-gap reference circuits this argument is irrelevant. There is even an advantage to the wide base of the parasitic pnp-transistor: a low ΔV_{be} mismatch in the order of $\sigma_{\Delta Vbe} = 0.2\,\text{mV}/\sqrt{\text{area}}$ where the area is measured in μm^2, see Sect. 5.3.7.

Resistors in a CMOS process are formed from diffused p- or n-type material.[7] In order to prevent large structures with lots of parasitics, unsalicidized material is us ed. This type of material will show resistivity values in the order of 50–150 $\Omega/\square$, mostly in combination with rather high temperature coefficients, see Table 5.3 in Sect. 5.2.3. For medium to high accuracy band-gap reference circuits, these temperature coefficients need to be part of the resistor model and need to be compensated by a slight shift of N_R, in order to achieve a correct result on silicon.

The operational amplifier needs a sufficiently high amplification factor in order to reduce the input offset and to provide a low-ohmic output. Two-stage Miller operational amplifiers in combination with a folded-cascode input stage will serve the purpose.

Figure 6.10 shows the input stage for the operational amplifier of a reference circuit. The input transistor pair consists of PMOS devices. An NMOS input stage (differential pair and current source) can be used as long as the NMOS threshold is low. The lowest input gate voltage on an NMOS stage equals $V_{T,N} + V_{drive} + V_{sat,current}$, the input transistor's threshold, and drive voltage plus the saturation voltage of the current source. The voltage over the pn-junction can reach values as low as 0.4 V at high temperatures, and although the temperature coefficient of the threshold voltage of NMOS devices is negative, the margins may become too small to keep high threshold NMOS devices plus the current source in inversion.

The input stage of Fig. 6.10 is build up as a folded-cascode stage. The NMOS current source $N2$ can be operated at a rather low saturation voltage ($V_D =$

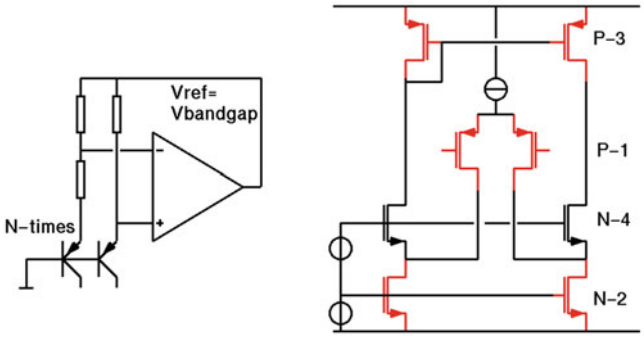


Fig. 6.10 The input stage of an operational amplifier for a band-gap reference circuit

[7] Specialized processes with layer of tungsten or titanium compounds offer resistors with far better properties.

0.2–0.3 V), whereas a current mirror would require an additional threshold voltage. The voltage over $N2$ determines how low the input pair $P1$ can be driven. If the transistors $P1$ have to remain in inversion, then $V_{G,P1} > V_D + V_{T,P1}$, where PMOS threshold is usually negative. This configuration allows low input voltages that may occur at high temperatures.

As indicated in Fig. 6.6 the input offset is multiplied by the resistive ratio to the output voltage. The random offset component will cause variations from one circuit to another. The random offset in a band-gap circuit is mainly caused by mismatch of the transistors in the operational amplifier. The transistor pairs P1, N2, and P3 contribute to the input referred mismatch. The cascode pair N4 is not relevant for the mismatch budget as in first order approximation this pair does not modify the current in the circuit.

Next to these hand calculations it is well possible to use circuit simulation for estimating the input referred mismatch voltage. Figure 6.11 shows a comparison of a typical simulation and a measured batch of reference circuits. Both the mean value and the standard deviation agree reasonably well for this sample size.

At higher temperatures the mismatch remains of equal importance.

Example 6.2. Calculate the input referred mismatch of the design in Fig. 6.10, when the transistor dimensions are P1: 90/5,N2: 28/7, and P3: 60/4, with $A_{VTn} = 14.2\,\text{mV}\mu\text{m}$, $A_{VTp} = 20.5\,\text{mV}\mu\text{m}$. In what way can the output variation of a group of band-gap circuits be reduced?

Solution. The input referred mismatch of pair P1 is easily found as it equals the threshold voltage mismatch of P1 using Eq. 5.16. The threshold voltage mismatches of pairs N2 and P3 are found in the same way, however, their input referred contribution needs to be calculated. The threshold voltage mismatch of N2 translates in a current mismatch via its transconductance. If the contribution of the other components is ignored, the input stage P1 will have to deliver an equal, but opposite

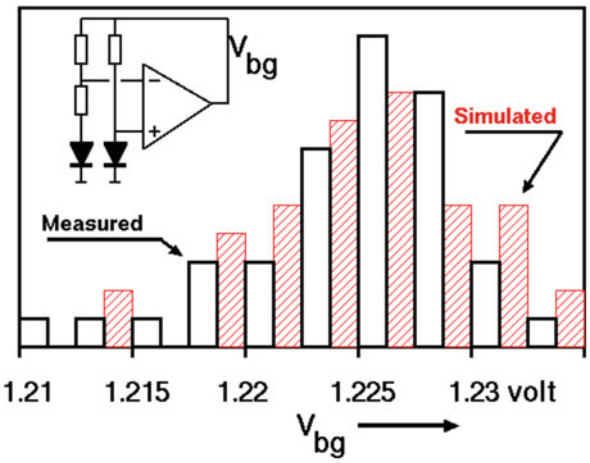


Fig. 6.11 Measured output voltage histogram of a band-gap reference circuit compared to simulation

Table 6.2 Calculation of input referred mismatch of the band-gap design, with $A_{VTn} = 14.2\,\text{mV}\mu\text{m}$, $A_{VTp} = 20.5\,\text{mV}\mu\text{m}$, $\mu_n = 2.5 \times \mu_p$

First design			Second design		
Transistor	σ_{VT}	σ_{Vin}	Transistor	σ_{VT}	σ_{Vin}
Dimensions	(mV)	(mV)	Dimensions	(mV)	(mV)
P1: 90/5	0.97	0.97	P1: 128/7	0.68	0.68
N2: 28/7	1.01	1.17	N2: 40/9	0.75	0.91
P3: 60/4	1.32	1.21	P3: 100/6	0.84	0.81
Total input referred:		1.94	Total input referred:		1.41
After multiplication (11x):		21.3 mV	After multiplication (7.8x):		11 mV
Measured reference s.d.:		24–27 mV	Measured reference s.d.:		9–10 mV

current to cancel the mismatch current from N2. In order to create this current an input referred voltage source in the gate connection of P1 is needed. This is the input referred mismatch voltage from N2: $\sigma_{Vin,N2}$. This voltage is found from equating the mismatch current from N2 to the input referred mismatch current in P1, via the transconductances of both devices:

$$i_{mismatch,N2} = g_{m,N2} \times \sigma_{VT,N2} = g_{m,P1} \times \sigma_{Vin,N2} \tag{6.7}$$

In a similar manner the contribution of P3 is referred back to the input. The total input referred mismatch is found in Table 6.2. The resulting variation in the output voltage of more than 20 mV is rather large. For that reason the circuit is modified to yield a lower spread of the reference voltage.

The first option is to change the dimensions of the transistors. Larger transistors will reduce the inherent mismatch. A gate-length reduction for P1 and longer transistors for N2 and P3 lead to a lower contribution of N2 and P3 due to a more favorable g_m ratio.

Next to that the multiplication factor of the opamp configuration $N_R + 1$ must be addressed. A lower resistor ratio is possible with a larger ΔV_{pn} value. This requires to increase the ratio of the pnp-transistors from 1:8 to 1:24 and to reduce $N_R + 1$ from 11 to 7.8. The second design is summarized on the right-hand side of Table 6.2.

Example 6.3. In Fig. 6.12 a very simple bias circuit is shown. All transistors operate in weak inversion mode. Calculate the current.

Solution. The current of an MOS transistor in weak inversion is

$$I_D = (W/L) I_0 e^{\frac{q(V_{GS}-V_T)}{mkT}} \tag{6.8}$$

Now current in the resistor is found in a similar way as with standard bipolar configurations. Assuming that the length of transistors M_1 and M_2 are equal:

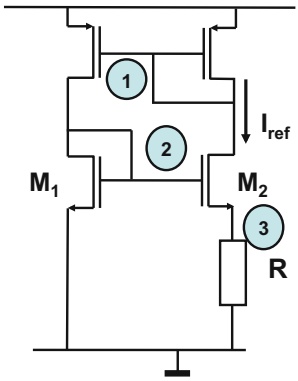


Fig. 6.12 This very simple bias circuit runs in weak inversion

$$I_{ref}R = V_{GS1} - V_{GS2} \;\Rightarrow\; I_{ref} = \frac{mkT}{qR}\ln\left(\frac{W_2}{W_1}\right)$$

It is clear that $W_2 > W_1$ to get a non-zero current. One of the often overlooked problems with this bias circuit is the need for a start-up method and the frequency stability. The circuit is in feedback mode with three poles as indicated by their numbers. If the poles are not separated (by one dominant pole), the circuit will behave as an oscillator.

6.2.5 Error Sources in Band-Gap Reference Sources

After examining the general properties of a band-gap circuit as in Fig. 6.2 some remarks are needed to analyze the most relevant error sources that affect the output voltage. The errors that contribute to variations of the output voltage from the designed value are split in two categories: some errors get multiplied by the resistor ratio of the opamp configuration and some result in common mode shifts.

The last category contains mainly two factors: the common variation of all resistors and of all diodes. Inspecting Eq. 6.4 shows that, e.g., dopant-level variation of all diodes affects the factor C_0. As a result, for example, 10 % more dope offsets the output reference voltage by $0.1kT/q \approx 2.5$–$3\,\text{mV}$. Similarly a change in the overall resistivity of the resistors will increase the current level in Eq. 6.4 from its intended value and result in a similar error.

More problematic are all errors that can be viewed as opamp input referred errors, as these are multiplied by the opamp-resistor configuration. The input referred random mismatch of the opamp transistors is the dominant factor. An input referred mismatch of σ_{Vin} yields an error in the output reference voltage $\sigma_{Vref} = (N_R+1)\sigma_{Vin}$.

A practical example could start from the mismatch parameter $A_{VT} = 4\,\text{mV}\mu\text{m}$. If the opamp input stage had three (more or less) equally contributing transistor

pairs, e.g., as in Fig. 6.10, with each an area of $10\,\mu m^2$, the resulting input referred mismatch is $\sigma_{Vin} = 2.2\,mV$. The variation in output voltage will be characterized by $\sigma_{Vref} = 20–25\,mV$.

Other factors are mostly less relevant. Resistor ratio errors contribute far below 1 %, so the effect on the output is less than $\sigma_{Vref} = 1–2\,mV$. Differences between the bipolar transistors will behave as opamp input errors; the relatively large area with an array of equal structures excludes lithography errors and will match well. A similar $\sigma_{Vref} = 1–2\,mV$ should be expected. The opamp gain is easily over 1000×, so its corresponding error and the variation in gain can be assumed negligible.

6.2.6 *Alternatives to Offset-Error Multiplication*

The dominant problem in the conventional band-gap circuit is the linear multiplication of all input referred errors in the generated PTAT voltage. An alternative to this multiplication is to build up the total PTAT voltage form a series connection of several PTAT sources, as shown in Fig. 6.13. Instead of a linear multiplication of the random offset error in one PTAT source now the independent contributions of N sources will add up to a multiplication factor for the random offset of $\sqrt{N}$ times. In order to realize the PTAT sources, mostly transistors biased in their weak inversion regime are used. This idea was followed in the design of a temperature compensated current source [114, 115].

A chopping scheme [van Veldhoven RHM, Dijkmans EC (1997) A bandgap reference using chopping for reduction of 1/f noise and random offset. Unpublished manuscript. Philips Research. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.2985&rep=rep1&type=pdf] [116] can eliminate the mismatch of the

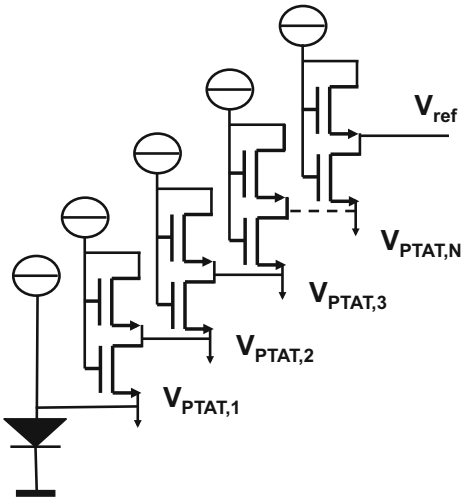


Fig. 6.13 A band-gap circuit with stacked PTAT sources [114, 115]

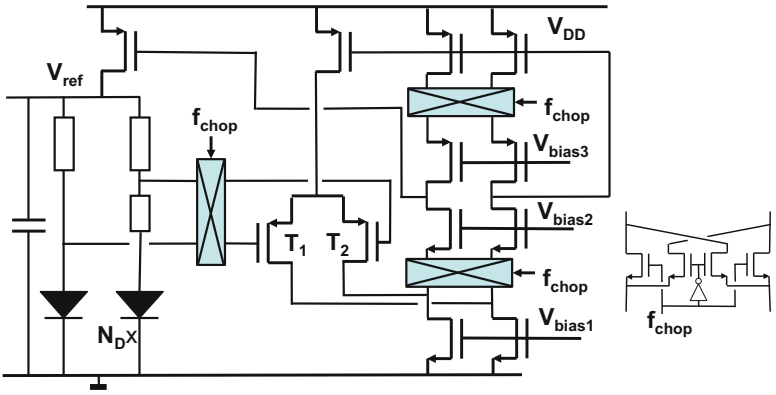


Fig. 6.14 A band-gap circuit with chopper [van Veldhoven RHM, Dijkmans EC (1997) A bandgap reference using chopping for reduction of 1/f noise and random offset. Unpublished manuscript. Philips Research. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.2985&rep=rep1&type=pdf] [116]. A start-up circuit is of course needed. van Veldhoven and Dijkmans (A bandgap reference using chopping for reduction of 1/f noise and random offset. Unpublished manuscript. Philips Research, 1997. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.2985&rep=rep1&type=pdf) uses $f_{chop} = 100\,\text{kHz}$ in a 0.5 µm CMOS process with 3 V power supply and $N_D = 24$

input stage of the input, Fig. 6.14. The opamp must then be designed differentially in order to use it in combination with chopping. So preferably only the input stage is chopped, not the output driver. The output voltage carries a ripple of $\pm\Delta V_{ref}$ at the chopping frequency and is only on-average correct. A high chopping frequency allows to use (switched) on-chip capacitors to remove this ripple, however, in a complex design the spurious components at multiples of the chopping frequency could upset RF parts. Moreover, the bandwidth of the chopped circuit should be sufficiently high, implying some power consumption. A low chopping frequency requires external filters. An unpublished paper[8] [van Veldhoven RHM, Dijkmans EC (1997) A bandgap reference using chopping for reduction of 1/f noise and random offset. Unpublished manuscript. Philips Research. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.2985&rep=rep1&type=pdf] reports a measured $\sigma_{ref} = 13\,\text{mV}$ before chopping and $\sigma_{ref} = 0.5\,\text{mV}$ in simulation with chopping. In [116] the authors reach a chopped output accuracy of $3\sigma_{Vref} = 0.75\,\%$, where a circuit without chopping as in Fig. 6.16 reaches $3\sigma_{Vref} = 1.7\,\%$.

[8] It is unknown how this paper reached the internet.

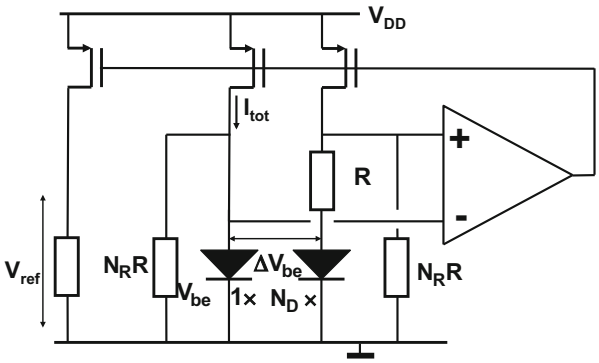


Fig. 6.15 A band-gap reference circuit for operation below 1.2 V [117]

6.2.7 Low-Voltage Band-Gap Circuits

Stacking a diode voltage and a temperature compensating voltage drop over a resistor requires a minimum power supply voltage of 1.4–1.5 V. In order to operate at lower supply voltages [117] presents the elegant solution shown in Fig. 6.15. In this circuit it is not the voltage that is made constant but the current. The PTAT current is generated in the diodes in exactly the same way as shown in the previous section. Adding to this PTAT current another PTAT current with an equally sized but opposite temperature coefficient, a temperature independent current can be obtained. Connecting two resistors over the diodes effectively creates this additional PTAT current that has a negative temperature coefficient. By choosing the proper N_R value,[9] the combination of the diode current and the resistor current allows to obtain a temperature independent current. A simple mirror operation defines any voltage level that is required. Here the absolute value of the resistor will show up in the output current and output voltage.

The catch in this circuit is that at extreme temperatures the current in either $N_R R$ or in the diodes consumes all current supplied in the feedback loop I_{tot}. The circuit will fail then. At very low temperatures with high V_{pn} and low power supply the current sources can be pushed out of saturation.

Another realization of a low-voltage band-gap reference circuit is achieved by feeding the PTAT current into a network formed by two resistors and a third diode Fig. 6.16. Proper tuning of the resistor values cancels the negative temperature coefficient of the diode at a level of, e.g., 800 mV. This circuit was also used to compare a PTAT voltage (controllable via a programmable resistor string) with the diode voltage. In this way a temperature sensor can be constructed. The measurements of the reference voltage as a function of the temperature show the characteristic second order curvature. Figure 6.17 shows a portion of the lay-out

[9] The same math applies here as for the standard circuit.

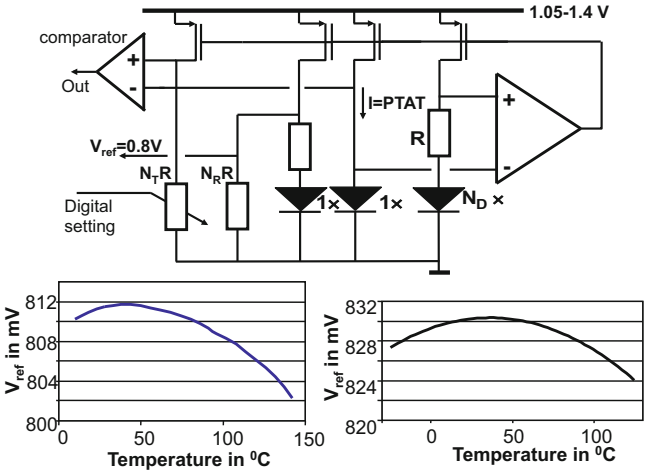


Fig. 6.16 A band-gap reference circuit of 0.8 V extended with temperature monitoring [118, 119]. *Lower left*: the output voltage of the circuit in 90-nm CMOS, $\sigma_{Vref} = 4.5$ mV over two batches. *Lower right*: the curve for the same circuit in 65-nm CMOS.

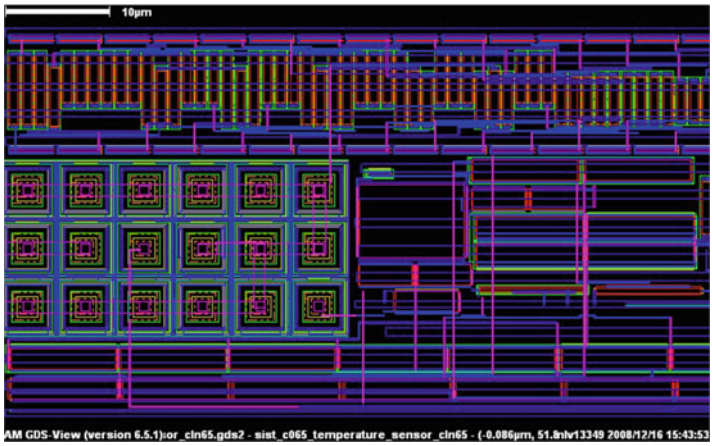


Fig. 6.17 The lay-out of a band-gap reference circuit in a 65-nm CMOS process as in Fig. 6.16

of this band-gap reference circuit. The common-centroid laid-out array of pnp-transistors and the large resistors and operational amplifier components are visible as well as the large opamp transistors.

Along similar lines as in [117–119] the circuit in Fig. 6.18 uses a DC-offset to mitigate headroom problems with low power supply voltages. The PTAT current source from Fig. 6.12 feeds into a standard resistor-diode circuit. Without further precautions the output voltage would show a flat temperature behavior at a level of 1.23–1.28 V. In [120] a DC-offset is introduced that will reduce the PTAT current

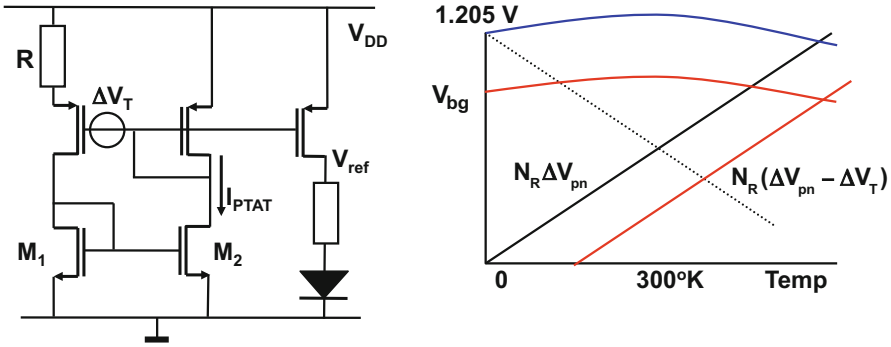


Fig. 6.18 In [120] the low-voltage operation of a band-gap reference is achieved by introducing a DC-offset

Table 6.3 A comparison of published band-gap references for low power supply

	Banba [117]	Cabrini [124]	Leung [123]	Annema [120]
Process	0.4 µm CMOS	0.13 µm CMOS	0.6 µm CMOS	0.16 µm CMOS
V_{ref}	0.515 V	0.798 V	0.603 V	0.944 V
σ_{Vref}	5.1 mV	13 mV		7.5 mV
V_{DD}	2.1 V	1.0 V	0.98 V	1.1 V
I_{DD}	2.2 µA	26 µA	18 µA	1.4 µA
Temp. range	27 °C–125 °C	−50 °C–160 °C	−0 °C–100 °C	−45 °C–135 °C
Area	0.1 mm^2	0.02 mm^2	0.24 mm^2	0.0025 mm^2

by a temperature independent amount of $\Delta V_T/R$. The output voltage will therefore be temperature independent at a value of $N_R \Delta V_T$ below the normal reference value. The DC-offset can be realized in various ways: in [120] the threshold modulation due to using different channel lengths is used.

In [121] the PTAT current is fed into a separate network. The results seem to indicate a larger spread. Table 6.3 lists some benchmarking design examples. A real avalanche of band-gap circuits for below-1 V operation is found in [122].

6.2.8 *Second Order Compensation*

The temperature dependence of the mobility factor in the diode current creates a dominant second order curvature in the output voltage of a reference circuit, as described in Eq. 6.6. This curve creates a deterministic deviation between 4 and 8 mV over a temperature range of −20 °C to 120 °C. Given the variations due to mismatch and other errors, this is typically not a dominant factor, but still many authors desire to compensate it.

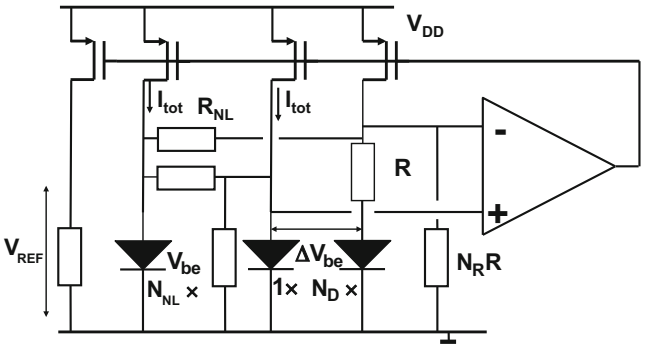


Fig. 6.19 Second order compensated band-gap reference circuit [125]

In [125] a simple method is proposed, depicted in Fig. 6.19. An additional group of N_{NL} diodes is fed with a constant current. The voltage over the resistors connecting the diodes for the PTAT current generation to the additional diode is

$$\Delta V_{NL} = V_{pn,1} - V_{pn,NL} \approx \frac{kT}{q} \ln\left(\frac{N_{NL} I_{PTAT}}{I_{tot}}\right) \tag{6.9}$$

where $I_{PTAT} \propto T$. The insight comes from the fact that the additional diode is fed with a temperature independent current causing the diode voltage to decrease linearly, while the voltage over the original diode is influenced by a PTAT current. Therefore the overall temperature dependence is identical to Eq. 6.5. Proper choice of the resistors R_{NL} cancels the second order behavior.

Example 6.4. A standard band-gap configuration has a diode ratio of 8 in a 0.18 µm technology. What is the minimum size of the input transistors of the opamp to keep the 1-sigma spread of the output voltage smaller than 1 %?

Solution. The diode ratio of 8 leads to a $\Delta V_{pn} = (kT/q)\ln(8) = 53\,\text{mV}$. With $V_{pn} \approx 0.7\,\text{V}$, the voltage over the resistor will be $V_{ref} - V_{pn} \approx 1.25 - 0.7 = 0.55\,\text{V}$. So the amplification ratio is in the range: $0.55/0.053 \approx 10$. The maximum allowed output spread is $\sigma_{Vref} = 0.01 \times V_{ref} \approx 12\,\text{mV}$. This σ_{Vref} is due to the input referred mismatch that is amplified by the opamp and resistor configuration. The input referred mismatch must therefore be less than $12\,\text{mV}/10 = 1.2\,\text{mV}$. As the A_{VT} parameter for 0.18 µm technology is in the range of 5 mVµm, the minimum area of the input transistors is found from:

$$\sigma_{\Delta VT} = \frac{A_{VT}}{WL} \rightarrow WL > 25\,\mu\text{m}^2$$

The topology of the opamp will determine whether the input referred mismatch is determined by the input stage or whether more components play a role.

6.3 Alternative References

6.3.1 DTMOST Reference Circuit

The sub-threshold current in an MOS transistor shows an exponential relation as a function of the gate–source voltage. An MOS transistor in weak inversion can be a viable replacement for the diffusion-well-substrate pnp device.

The injection of charge from the source into channel is identical t o a normal pn-junction, see Fig. 6.20. Unfortunately only the source side of this junction can be directly accessed. The potential in the channel is formed by the electrostatic balance between gate potential and local substrate voltage. In case the gate potential is changed with respect to the other terminals, a capacitive division between gate capacitor and substrate capacitor will determine this channel potential. This division is represented by the factor $m \approx 1.1 \ldots 1.3$ in the exponent of Eq. 6.8. The weak inversion current generated by an increasing gate voltage will follow a 70–80 mV/decade current slope, as shown in Fig. 6.21. Huang et al. [126] gives a sub-threshold based reference with switched-capacitor amplification.

On the other hand, also the source potential can be varied with respect to the gate and substrate voltages. Now there is no voltage division and a perfect diode curve is found, see Fig. 6.21. This idea is followed in a DTMOS transistor [115], where gate and local substrate are tied together. It is also very well possible to have the gate and local substrate connected to different DC potentials.

It is therefore possible to replace the diode in a reference circuit by an MOS transistor with a 60 mV/decade slope, see Fig. 6.22. However, in an absolute sense, the threshold voltage will still be part of the total voltage. This causes sensitivity to global and local variations and to specific temperature dependencies of the work functions.

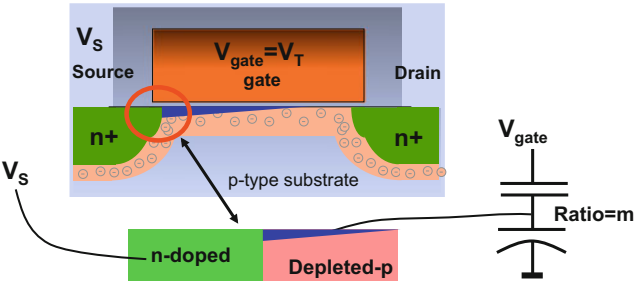


Fig. 6.20 An MOS transistor consists of a diode between the source and the substrate that pulls the emitted electrons in the drift region towards the drain

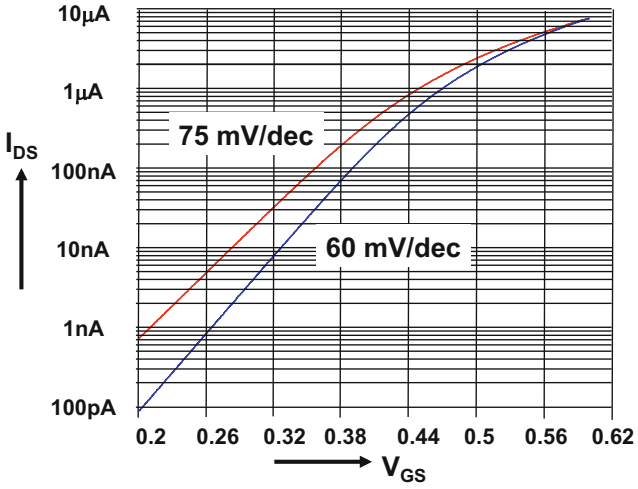


Fig. 6.21 A 4/1 µm PMOS transistor in 65-nm CMOS is operated in the weak inversion regime. The *upper curve* shows the current for an increasing gate voltage, the *lower curve* for a decreasing source voltage

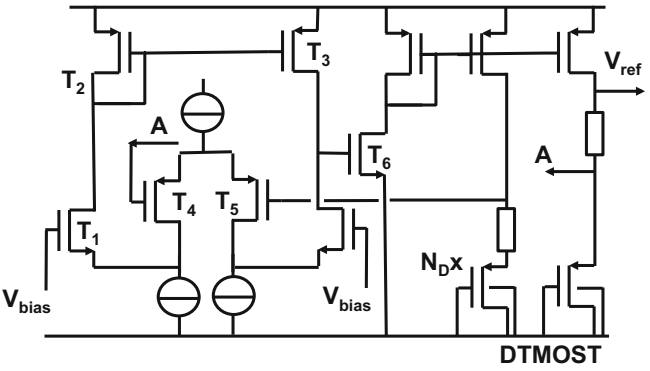


Fig. 6.22 A reference circuit using bulk-connected DTMOS transistors [115]. The two *right-most* branches form the diode pair. On the *left side* a folded-cascode amplifier

6.3.2 Technological Options

The band-gap of silicon is a stable and reliable physical quantity to serve as the base for a reference voltage. Some other quantities have also been proposed.

- The threshold voltage of the MOS transistor is a potential candidate for designing a reference source. Running a current in a diode-connected transistor gives

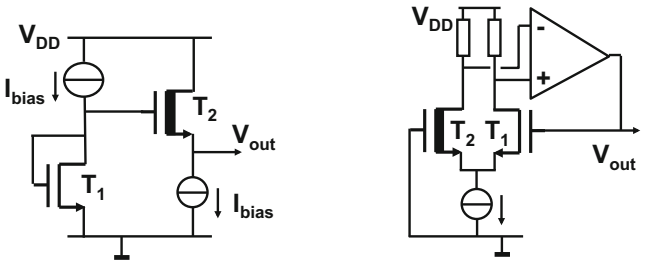


Fig. 6.23 Two ways of generating the difference between two threshold voltages of T_1 and T_2

$$V_{ref} = V_T + \sqrt{\frac{2I}{\mu C_{ox} W/L}} \tag{6.10}$$

Both the threshold of the MOS transistor and the mobility show a considerably temperature coefficient. The NMOS threshold decreases with 1–3 mV/°C, while the square-root term increases in value due to the decreasing mobility. There is a sweet spot where both effects compensate. Industrially this is considered to be not a reliable solution. Therefore some form of controlled compensation is needed, which is normally applied by making the current temperature dependent [11].

- The difference between two threshold voltages can be exploited in various ways. In the circuit of Fig. 6.23 (left) the W/L ratios and currents are equal. Now the output voltage is easily found by subtracting Eq. 6.10 for transistor T_2 from the equation for T_1. In first order the current dependent part will cancel and $V_{out} = V_{T,T1} - V_{T,T2}$. If T_1 and T_2 have threshold voltages of the same polarity V_{out} is small and will suffer a lot from variations. This circuit is more interesting if opposite threshold voltages are available [127, 128].
- The threshold voltage of an EEPROM device can be trimmed during the programming process. This method and derived ideas require, however, some form of factory trimming.[10]
- In Finfet technology an alternative is needed as there are no parasitic pnp-structures available. In [129] a lateral p-n junction covered with a polysilicon line (known as lubistor) is used. The results are promising, but further research is needed.

Despite some interesting attempts these principles have never gained much acceptance compared to band-gap based references.

Example 6.5. In Fig. 6.24 a standard four-transistor reference circuit is shown, however, one circuit is erroneously connected. Which one?

[10]This option is mentioned in product announcements by Xicor and Intersil, e.g., application note AN177.

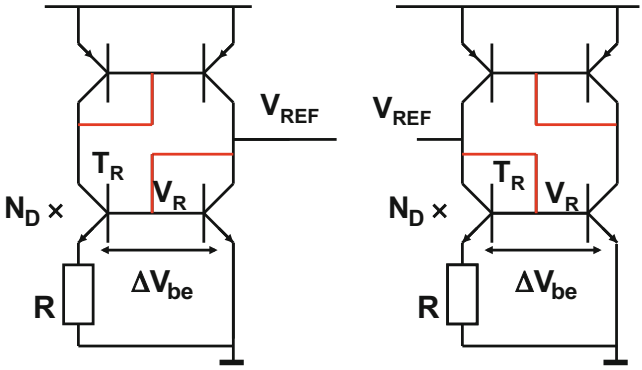


Fig. 6.24 The most often made mistake in the 4-transistor cell is the position of the collector-base connections. Which diagram is correct: left or right?

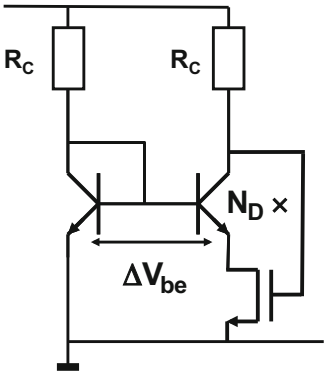


Fig. 6.25 An MOS transistor with feedback replaces the resistor

Solution. Push the common npn-base node V_R on both circuits to a lower value. This voltage change will cause in the right-hand npn transistor a huge loss of current as the voltage change affects the current via an exponential function. The left-hand current will also go down but less, as the voltage drop will fall mostly over the resistor, with a linear decrease of current. So the npn-current in the left branch becomes larger than in the right branch, which is the case in both circuits.

In the left-hand circuit, the larger current is mirrored via the pnp-transistors to the right branch and starts pulling up node V_R: this counteracts the original excursion and means that the left-hand circuit is in negative feedback.

In the right-hand circuit, the smaller current is mirrored and will allow the node V_R to be pushed down easier: positive feedback which renders the circuit unstable.

Example 6.6. Figure 6.25 uses an MOS transistor in its linear regime as a resistor. What are the (dis)advantages?

Solution. For very low currents high value resistors are needed with associated area penalty. This solution not only combines an MOS transistor as a resistor but also allows some feedback, to reduce noise amplification.

Exercises

6.1. A process allows to design resistors with a global spread of $\pm 20\,\%$. What will be the impact of this spread on the reproducibility of a standard band-gap circuit?

6.2. What bandwidth is advisable for the opamp in a standard band-gap circuit of Fig. 6.1 if this band-gap has to feed a 25 Ms/s switched-capacitor digital-to-analog converter?

6.3. Discuss other solutions to the above problem: decoupling, a buffer circuit, multiple references.

6.4. A start-up method in Fig. 6.9 uses a diode between the pnp-transistor and the npn-transistor. Discuss the advantages and disadvantages.

6.5. Propose a start-up method for the circuit in Fig. 6.2 other than indicated in the text.

6.6. The total area available for the opamp in a CMOS process with $A_{VT} = 4\,\mathrm{mV\mu m}$ is $40{\times}40\,\mu\mathrm{m}^2$. Propose an opamp topology and give an estimate for the resulting reproducibly of the band-gap voltage.

6.7. Estimate for all band-gap circuits in this chapter, what the power supply requirements are.

6.8. Setup the calculation for N_R in the circuit shown in Fig. 6.15.

Chapter 7
Digital-to-Analog Conversion

Digital-to-analog converters fulfill two important roles in the data conversion chain. A digital-to-analog converter is needed at the end of the chain for converting the digital signal back into the physical domain. Next to that every analog-to-digital converter needs some form of digital-to-analog conversion for its operation. These two functions directly influence the requirements posed on the digital-to-analog conversion.

A digital-to-analog converter that has to deliver a signal to the physical world acts in the continuous time domain and the signal has to show a high quality at every time moment. Moreover the signal must be delivered at some power level to a load impedance.

In an analog-to-digital converter the value delivered by the digital-to-analog converter is relevant only at a few (perhaps only one) time moments. The performance on other time moments is not critical. Together with a minimum demand on the drive capabilities, the demands of this application on the converter are mostly much less challenging.

The application constraints limit the freedom of the choice of the architecture and of the physical domain. In this chapter, the architectural and physical domain options are analyzed. Then some realizations of digital-to-analog converters per domain illustrate the combination of these aspects.

M. Pelgrom, *Analog-to-Digital Conversion*, DOI 10.1007/978-3-319-44971-5_7

Table 7.1 Various forms of digital representation

Straight binary		Two's complement		Sign+magnitude		Gray coded	
15	1111	7	0111	7	0111	15	1000
14	1110	6	0110	6	0110	14	1001
13	1101	5	0101	5	0101	13	1011
12	1100	4	0100	4	0100	12	1010
11	1011	3	0011	3	0011	11	1110
10	1010	2	0010	2	0010	10	1111
9	1001	1	0001	1	0001	9	1101
8	1000	0	0000	0	0000	8	1100
7	0111	−1	1111	0	1000	7	0100
6	0110	−2	1110	−1	1001	6	0101
5	0101	−3	1101	−2	1010	5	0111
4	0100	−4	1100	−3	1011	4	0110
3	0011	−5	1011	−4	1100	3	0010
2	0010	−6	1010	−5	1101	2	0011
1	0001	−7	1001	−6	1110	1	0001
0	0000	−8	1000	−7	1111	0	0000

7.1 Representations

7.1.1 Digital Representation

The input to a digital-to-analog converter is a digital sample. Table 7.1 shows various representations of digital signals.[1] The most simple approach is to assume a positive signal. The straight-binary code in the first column is well suited for positive signals. Of course most systems use signals with negative values as well. There are several ways to represent negative signals in a conversion process. The choice how to represent the signal will influence several aspects of the conversion and of the analog and digital processing. Negative signals can be used if the entire scale is shifted by half of the full amplitude. In this "two's-complement" representation, the mid-level value is the signal zero. In the digital domain now the positive and negative signals can be properly handled. Addition and subtraction can be executed without prior knowledge of the signs of the operands. Multiplication requires the digital numbers to be extended to fit a format corresponding to the result. Positive numbers are extended by zeros and negative values with ones. A direct conversion of a "two's complement" code in the analog domain requires a positive and negative analog power supply. This is mostly not an economical solution, therefore the code "0000" corresponds in the analog domain with half of the reference value or half

[1] Many more representations exist, this table only lists the ones used in this book.

of the power supply voltage. Now small signals in the analog domain move around half of the reference value. This leads often to power and noise performance loss. Other artifacts associated with half of the reference value will further deteriorate the signal-to-noise ratio. For example, in current digital-to-analog converters, half of the reference value corresponds to half of the maximum current. Thereby a zero-valued signal will show thermal and 1/f noise which is associated with this current. Obtaining a good signal-to-noise ratio for small signals is made difficult here by the choice for two's complement representation.

The "sign and magnitude" code is linked to the design style in analog hardware. The MSB of the code is the sign and the remaining part of the code is the amplitude of the signal. From a circuit design point of view the MSB signal can directly be used to switch, e.g. a polarity switch in a digital-to-analog converter. The amplitude component of a "sign and magnitude" code is straight binary. Consequently this kind of code allows implementations that avoid problems with noisy half-reference values. In the digital domain this code is less pleasant: a digital decoder is needed before this signal can be handled by the software in digital signal processors.

If a "sign-and-magnitude" signal is rounded or truncated in the digital domain with simple rounding or truncation rules an error will occur, see Fig. 7.1. In the case of rounding or truncation for a "straight binary" or "two's-complement" signal truncations of positive and negative numbers will result in a shift in the same direction. For "sign-and-magnitude" signals truncation of the positive and the negative part of the signal reduces the amplitude and shifts both sides towards zero. Such straight-forward rounding or truncation will give a cross-over problem near zero and a distortion component.

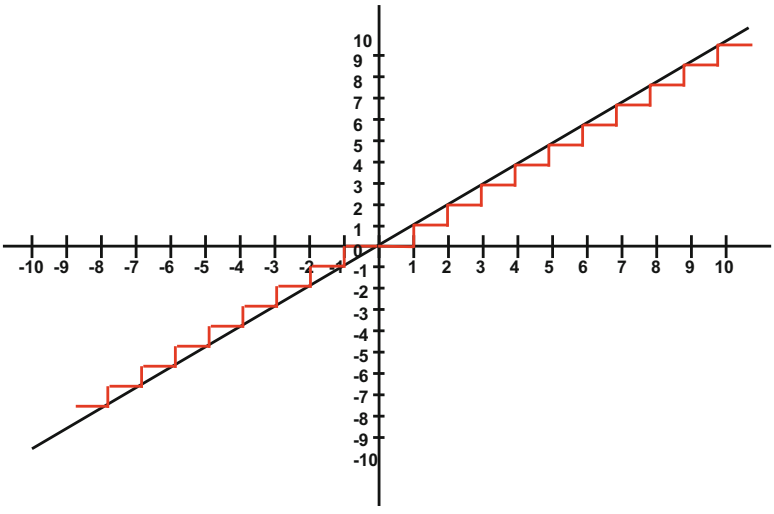


Fig. 7.1 Basic truncation (or rounding) creates distortion near the zero-code

Non-linear operations in the digital domain must always be closely examined irrespective of the choice for a digital representation.

7.1.2 Unary Representation

A reference quantity forms the basis for the digital-to-analog conversion. The next step is the subdivision of this reference in fractions that allow the generation of a quantity corresponding to an LSB or multiples of an LSB. Of course it is also possible to generate directly a reference of the size of an LSB and multiply this value. However, this will also multiply the noise and interference. Starting with a full-scale reference value reduces these sensitivities and results in the best performance.

The two most commonly used techniques for combining elementary units created by a reference are the unary and binary representation. Figure 7.2 shows both architectures.

The unary representation uses a series of 2^N identical elements (elements are resistors, capacitors, currents, devices, etc.). A unary numerical value is created as

$$B_u = \sum_{i=0}^{i=2^N-1} b_i = b_0 + b_1 + b_2 \ldots + b_{2^N-1} \tag{7.1}$$

where each coefficient b_i equals either "0" or "1." Some authors use the terms "thermometer" code for this scheme, due to the obvious resemblance with temperature measurement. The analog equivalent is formed by summing copies of the physical equivalent A_{LSB} of an LSB:

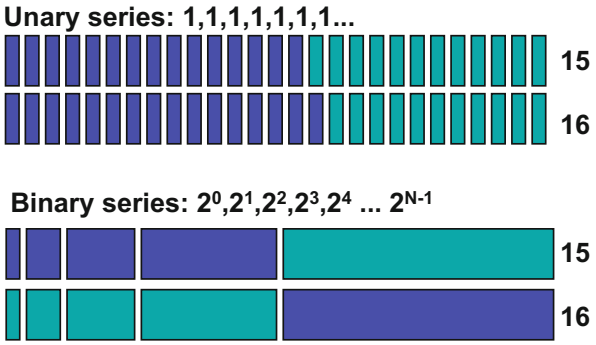


Fig. 7.2 Two basic techniques for digital–analog conversion: unary and binary representation. For both representations two signals are shown corresponding to a value of 15 and 16 LSBs

$$A = \sum_{i=0}^{i=2^N-1} b_i A_{LSB,i} = b_0 A_{LSB,0} + b_1 A_{LSB,1} + b_2 A_{LSB,2} \ldots + b_{2^N-1} A_{LSB,2^N-1} \quad (7.2)$$

A 1 LSB higher value can easily be created by adding one new element to the previously selected elements, in Fig. 7.2 (upper) the code 15 is increased to 16. The obvious advantage of this method is that it provides an absolute guarantee on monotonicity (see Fig. 4.10). A practical implementation will consist of 2^N elements (resistors, capacitors, or current sources) attached to an extensive switching matrix. Every element needs a switch. Also all dynamic events (e.g., switch charge dump) in a unary converter will be proportional to the code, which in first order cancels the effect. A converter based on unary coding will grow exponentially in area with N. Till $N = 10 \ldots 12$ unary coded converters will result in a good, and economically usable solution. This technique can be applied to resistor strings, current source arrays, capacitor arrays, and in timing (counting conversion).

7.1.3 Binary Representation

In order to avoid the exponential growth of components in a unary architecture, the exponential behavior must be included in the representation itself. In a binary structure the elements are chosen such that the resulting voltages or currents form an exponential series.

$$B_b = \sum_{i=0}^{i=N-1} b_i 2^i = b_0 + b_1 2^1 + b_2 2^2 \ldots + b_{N-1} 2^{N-1} \quad (7.3)$$

or in the physical domain:

$$A = \sum_{i=0}^{i=N-1} b_i A_{LSB+i} = b_0 A_{LSB} + b_1 A_{LSB+1} + b_2 \ldots + b_{N-1} A_{MSB} \quad (7.4)$$

As a simple switch has two positions, therefore it is practical (but not necessary) to choose 2 as a base. The example in Fig. 7.2 shows a series of elements with the values: $A_{LSB}, A_{LSB+1}, \ldots A_{MSB} = 1, 2, 4, 8$, and 16. A binary coded converter switches the weighted elements on and off. From 14 (code 01110) to 15 (01111) the lowest weighted element is added. In the example of Fig. 7.2 (lower) the dark colored elements add up to the value of 15. After another increment, all elements chosen up to then must be switched off and the element for the value 16 is switched on. The implicit disadvantage of this method is the transition from one value to another value, at codes where many bits flip (e.g., 01111→10000). Although most transitions will result in an almost perfect LSB change, the transitions in the higher bits will cause a high-valued element to switch on and all other elements to switch

0	-
1	Δ_0
11	$\Delta_0+\Delta_1$
111	$\Delta_0+\Delta_1+\Delta_2$
1111	$\Delta_0+\Delta_1+\Delta_2+\Delta_3$
11111	$\Delta_0+...+\Delta_4$
111111	$\Delta_0+...+\Delta_5$
1111111	$\Delta_0+...+\Delta_6$
11111111	$\Delta_0+...+\Delta_6+\Delta_7$
111111111	$\Delta_0+...+\Delta_8$
1111111111	$\Delta_0+...+\Delta_9$
11111111111	$\Delta_0+...+\Delta_{10}$
111111111111	$\Delta_0+...+\Delta_{11}$
1111111111111	$\Delta_0+...+\Delta_{12}$
11111111111111	$\Delta_0+...+\Delta_{13}$
111111111111111	$\Delta_0+...+\Delta_{14}$

0000	-
0001	Δ_0
0010	Δ_1
0011	$\Delta_0+\Delta_1$
0100	Δ_2
0101	$\Delta_0+\Delta_2$
0110	$\Delta_1+\Delta_2$
0111	$\Delta_0+\Delta_1+\Delta_2$
1000	Δ_3
1001	$\Delta_0+\Delta_3$
1010	$\Delta_1+\Delta_3$
1011	$\Delta_0+\Delta_1+\Delta_3$
1100	$\Delta_2+\Delta_3$
1101	$\Delta_0+\Delta_2+\Delta_3$
1110	$\Delta_1+\Delta_2+\Delta_3$
1111	$\Delta_0+\Delta_1+\Delta_2+\Delta_3$

Fig. 7.3 *Left*: the unary code is increasing and at every step one new error is added. *Right*: in the binary code there is a switch-over of elements at major transitions, with more impact of the errors

Table 7.2 The unary and the binary architectures are compared on a number of aspects

Architecture	Unary	Binary
Monotonicity	By design	Not guaranteed
Number of elements	$\propto 2^N$	$\propto N$
Area, parasitics	$\propto 2^N$	$\propto N$
INL systematic	Gradient	Small
INL random	$\propto \sigma_{element}\sqrt{2^{N-1}}$	Same with DNL errors
DNL random	$\propto \sigma_{element}$	$\propto \sigma_{element}\sqrt{N,\ldots,2^{N-1}}$ see specific section
Switching energy	$\propto$ Signal	Can be large and non-linear
Decoder complexity	2^N-to-1	Simple
Power	Choice for R, I, or C implementation dominates power	
Noise	Similar for similar power levels	

off. Both values should differ the physical equivalent of one LSB, however errors in the physical quantities $A_{LSB}, A_{LSB+1}, \ldots$ can easily create deviations. Figure 7.3 shows in a tabular mode the difference between the unary and binary architectures. If a higher exponent element is smaller than the sum of the lower exponential elements, a non-monotonicity in the transfer curve will arise. In this case an increment on the digital input code will result in a decrementing analog output. In control loops instability around such a non-monotonicity will upset the system.

Table 7.2 summarizes the main differences.

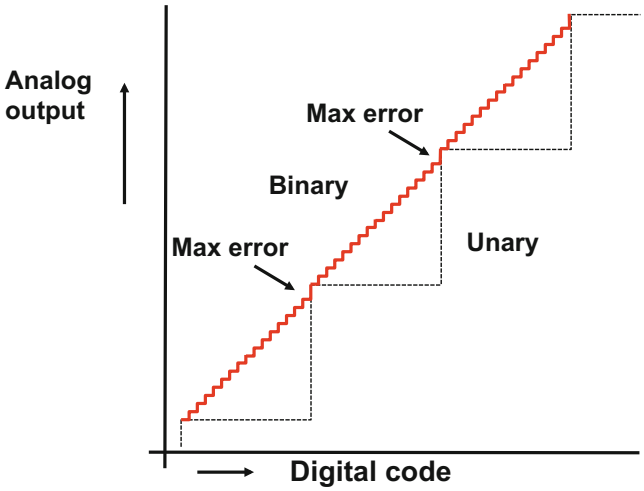


Fig. 7.4 In a segmented architecture the coarse unary steps are detailed with a section of binary elements. In this example the binary section counts four bits

7.1.4 Segmentation

The unary and the binary architecture differ in many aspects, see Table 7.2. In high resolution ($N > 8$) converters the better DNL performance of the unary architecture must be traded off against the smaller size of the binary architecture. In all implementation domains (R,I,C) solutions exist that combine the best of both. These "segmented" digital-to-analog converters realize N_{MSB} bit resolution with a unary architecture and combine that with an N_{LSB} bit binary structure. Figure 7.4 shows an example where the coarse unary steps are supplemented by a binary addition. The unary steps must have the accuracy of an LSB, which is the smallest step of the binary structure. As the same binary structure is used for every coarse unary step, a deviation in the binary pattern will repeat over all unary sections.

7.1.5 Unary: INL and DNL

The integral non-linearity (INL) of a digital-to-analog converter is determined by systematic and statistical deviations. Systematic or deterministic errors are due to gradients, distortion as described in Sect. 4.5, or inaccuracies in the lay-out, see Chap. 5. Deterministic errors can be eliminated or reduced and often statistical deviations are the remaining reason for INL and DNL errors. Statistical deviations can be minimized, but require area and/or power.[2]

[2]In this section the effect of statistical errors is examined. In the following chapters compensation and calibration techniques are discussed.

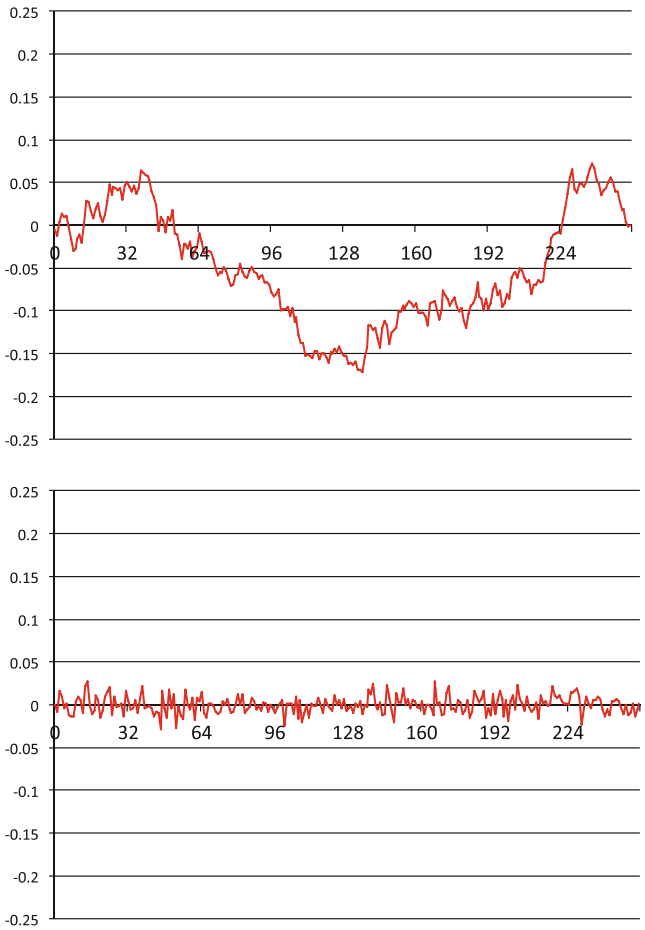


Fig. 7.5 This simulated 8-bit unary digital-to-analog converter is constructed of 255 elements with each a nominal value LSB and with an independent standard deviation of 1 % LSB. The INL and DNL curves show rather smooth curves as in this architecture little deviations caused by added at each code

In a unary architecture every next code is generated by adding a new element to the group of previously selected elements. Figure 7.5 shows the typical behavior of the INL and DNL curves for this unary architecture. The DNL is simply determined by the variation in each element from its ideal LSB value.

The INL can take many different shapes, but mostly some pattern is visible as in Fig. 7.5, causing distortion products.

Figure 7.6 shows the simulation of a unary converter of 256 elements with a relative current mismatch of 1 %. The coding of the current sources is in a thermometer manner: each sample is built up starting with source number 1 until the number of elements corresponding to the digital code is active. If the random

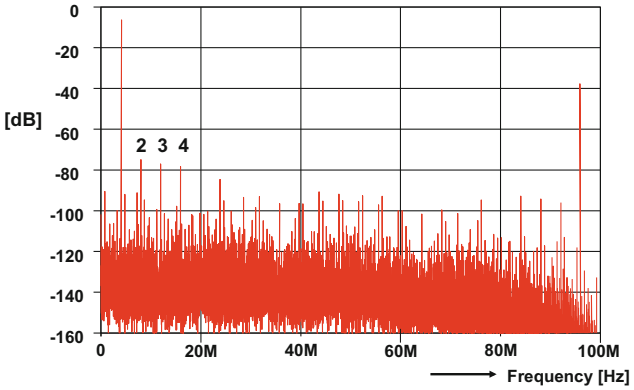


Fig. 7.6 Spectrum of a digital-to-analog converter with random mismatch. An array of 256 elements (in this case: current sources) is simulated with a normally distributed mismatch of $\sigma_I/I = 1\,\%$. The signal frequency is 3.97 MHz at a 100 Ms/s sample rate. The second, third, and fourth order distortion components are labeled

process creates an error in an element, then this will result in an error at a fixed value of the signal. The random errors in the elements create an INL curve as in Fig. 7.5. The conversion of a sine wave signal leads partially to distortion products of the signal and partially in a fixed-pattern noise. Data-weighted averaging techniques circumvent the occurrence of harmonic distortion due to random errors, see Sect. 7.6.3.

If the physical equivalent of an LSB is A_{LSB}, then the population of 2^N elements in a unary architecture is characterized by a statistical distribution. Most elements (resistors, capacitors, transistors, currents, charges, etc.) consist of a large sum of many independent events, e.g. the value of a resistor/conductance is composed of many small, atomistic conductivities. For large sums of independent events with identical (but unknown) probability distributions, the Central Limit Theorem dictates that the sum behaves as a Gaussian distribution. This assumption is implicit in most analyses.

For the m-th element $A_{LSB}(m)$ the mean and standard deviation of the Gaussian distribution for $A_{LSB}(m)$ are

$$E[A_{LSB}(m)] = E\left[2^{-N}\sum_{i=0}^{2^N-1} A_{LSB}(i)\right] = A_{LSB} \tag{7.5}$$

$$\sigma(A_{LSB}(m)) = \sigma_{A_{LSB}} \quad \rightarrow \frac{\sigma(A_{LSB}(m))}{A_{LSB}} = \frac{\sigma_{A_{LSB}}}{A_{LSB}} \tag{7.6}$$

where m is running from $m = 0, 1, \ldots, (2^N - 1)$ and every element $A_{LSB}(m)$ is taken from the same statistical distribution. The last term represents the relative standard deviation. Note that although the physical equivalent of an LSB equals A_{LSB} the LSB

for a particular converter is the sum of the elements divided by 2^N, which is close to but not necessarily exactly equal to A_{LSB}. This complicates the next derivation as the elements that are not used for the signal still influence the outcome as they too determine the range over which the INL is measured and the physical value of the LSB.

Now the DNL for any position m is an event from the distribution with mean and standard deviation:

$$E[DNL(m)] = E\left[\frac{A_{LSB}(m)}{A_{LSB}} - 1\right] = 0, \qquad \sigma_{DNL}(m) = \frac{\sigma_{A_{LSB}}}{A_{LSB}} \tag{7.7}$$

Figure 7.5 (right) shows the DNL curve for $\sigma_{A_{LSB}}/A_{LSB} = 1\,\%$. For this group of 255 elements maximum deviations up to $3\sigma_{DNL} = 0.03\,\text{LSB}$ are observed. The maximum DNL is determined by the largest deviating element amongst 2^N random varying elements with each a relative spread of σ_{DNL}. In order to guarantee the DNL for a production lot the estimation will have to be extended to P converters times 2^N elements per converter.

In unary architectures the effect of statistical errors is influencing specifically the INL performance. A group of m elements with value y_1 is forming the signal, and there is a group $(2^N - m)$ elements with value y_2 left over:

$$y_1 = \sum_{i=0}^{m-1} A_{LSB}(i), \text{ with } E(y_1) = mA_{LSB}, \sigma_{y1} = \sigma_{A_{LSB}}\sqrt{m} \tag{7.8}$$

$$y_2 = \sum_{i=m}^{2^N-1} A_{LSB}(i), \text{ with } E(y_2) = (2^N - m)A_{LSB}, \sigma_{y2} = \sigma_{A_{LSB}}\sqrt{2^N - m} \tag{7.9}$$

The INL at position m is now:

$$INL(m) = \frac{2^N \sum_{i=0}^{m-1} A_{LSB}(i)}{\sum_{i=0}^{2^N-1} A_{LSB}(i)} - m = \frac{2^N y_1}{y_1 + y_2} - m \tag{7.10}$$

The variance of the INL is now found by applying the statistics formula for multiple stochastic variables, Eq. 5.11:

$$\begin{aligned}
\sigma^2_{INL}(m) &= \left(\frac{\partial INL(m)}{\partial y_1}\right)^2 \sigma^2_{y1} + \left(\frac{\partial INL(m)}{\partial y_2}\right)^2 \sigma^2_{y2} \\
&= \left(\frac{y_2}{(y_1 + y_2)^2}\right)^2 m\sigma^2_{A_{LSB}} + \left(\frac{-y_1}{(y_1 + y_2)^2}\right)^2 (2^N - m)\sigma^2_{A_{LSB}} \\
&= \frac{m(2^N - m)}{2^N} \frac{\sigma^2_{A_{LSB}}}{A^2_{LSB}}
\end{aligned} \tag{7.11}$$

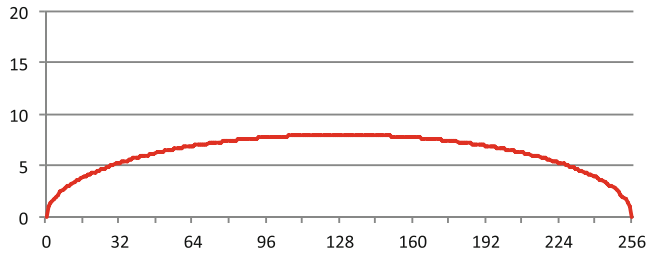


Fig. 7.7 The function $\sqrt{m(2^N - m)2^{-N}}$ shows the typical envelop of many random INLs

In Fig. 7.7 a plot is shown of the function $\sqrt{m(2^N - m)2^{-N}}$ that forms the heart of Eq. 7.11. The maximum value of the variance occurs in the middle at $m = 2^{N-1}$ for which the variance is found as

$$\sigma^2_{INL}(m = 2^{N-1}) = 2^{N-2}\frac{\sigma^2_{A_{LSB}}}{A^2_{LSB}} \tag{7.12}$$

The position of the maximum value of the INL error in an arbitrary unary converter is not necessarily found in the middle of the structure, although there the probability is the highest. Nevertheless Eq. 7.12 can serve to estimate the INL. With the help of Table 2 the sigma margin can be chosen.

7.1.6 Binary: INL and DNL

Figure 7.8 shows a typical integral and differential nonlinearity curve for a binary digital-to-analog converter. The jumps at code 64, 128, and 192 indicate that the difference between the elements representing the MSB and MSB-1 and the sum of the elements for the lower bits deviates severely from 1 A_{LSB}. The error occurring at code 64 is identical to the error at code 192, because these are caused by the same switching sequence only differing by the MSB. The same observation holds for the errors at the MSB-3 in this example. The repetition of the pattern consisting of the LSB, LSB+1,.. is another trademark of a binary architecture.

The major DNL error will occur at the transition of the 2^{N-1} elements forming the MSB section to the remaining $(2^{N-1} - 1)$ elements for the remaining bits. Its variance can be estimated[3]:

$$\sigma^2_{\text{DNL MSB}} = \frac{\sigma^2_{A_{MSB}} + \sigma^2_{A_{rest}}}{A^2_{LSB}} = \frac{2^{N-1}\sigma^2_{A_{LSB}}}{A^2_{LSB}} + \frac{(2^{N-1} - 1)\sigma^2_{A_{LSB}}}{A^2_{LSB}} = (2^N - 1)\frac{\sigma^2_{A_{LSB}}}{A^2_{LSB}}$$

[3]Formally the more correct approach involves starting with $DNL = 2^N(y_1 - y_2)/(y_1 + y_2) - 1$.

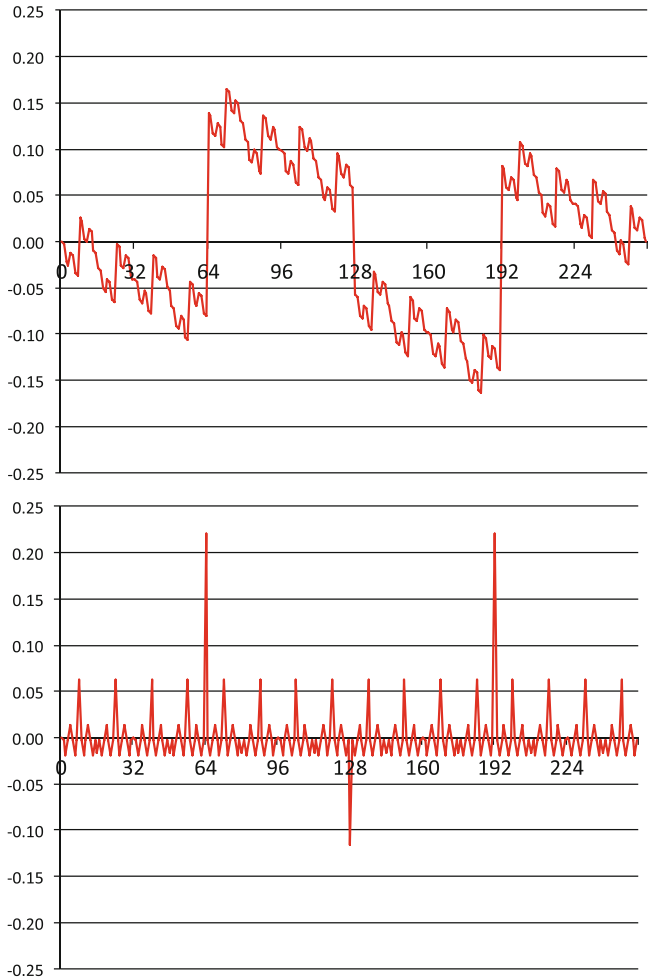


Fig. 7.8 The 255 elements of Fig. 7.5 are wired to form a binary architecture. Each unit element has a nominal value of 1 LSB and a standard deviation of 1 %. The elements are grouped together to form a binary sequence: 1,2, 4, 8, . . . , 128. The INL and DNL curves of the 8-bit binary digital-to-analog converter show the typical jumps at the MSB transition, the MSB-1 transition, etc.

Similarly:

$$\sigma_{\text{DNL MSB}} = \sqrt{(2^N - 1)}\frac{\sigma_{A_{LSB}}}{A_{LSB}} \tag{7.13}$$

$$\sigma_{\text{DNL MSB}-1} = \sqrt{(2^{N-1} - 1)}\frac{\sigma_{A_{LSB}}}{A_{LSB}} \tag{7.14}$$

$$\sigma_{\text{DNL MSB}-2} = \sqrt{(2^{N-2} - 1)}\frac{\sigma_{A_{LSB}}}{A_{LSB}} \tag{7.15}$$

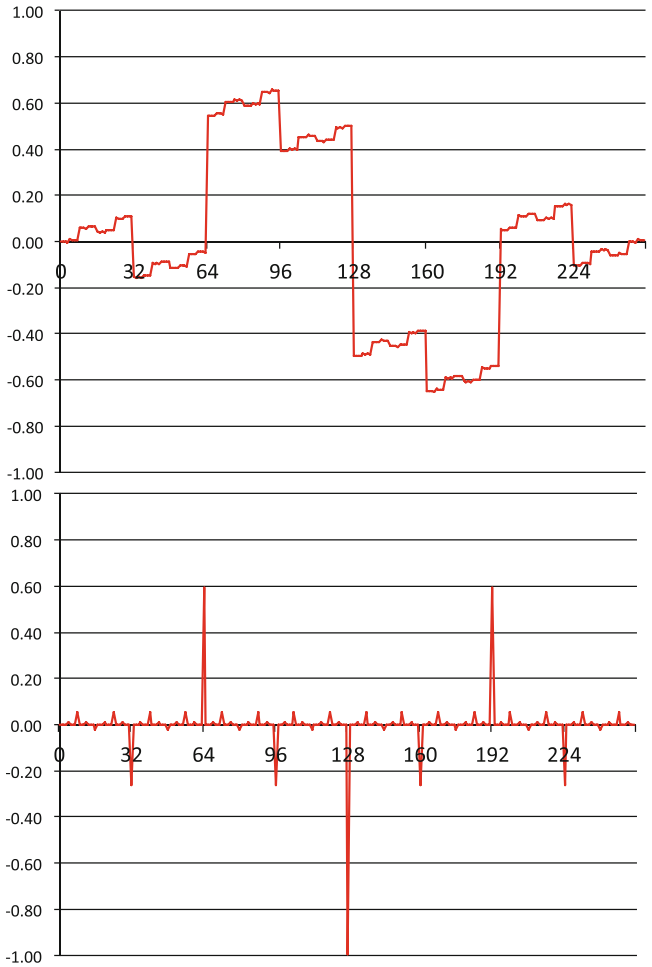


Fig. 7.9 This simulation uses eight elements scaled as an exponential series with nominal values: 1, 2, 4, …, 128. Each element (irrespective of its nominal value) has an independent standard deviation of 1 % of its nominal value. Note that the vertical scale is 4× the scales of the other figures in this section

The DNL MSB-1 error occurs three times: at codes 2^{N-2}, $2 \times 2^{N-2}$, and $3 \times 2^{N-2}$. The middle transition is part of the DNL MSB error.

The curves for the design in Fig. 7.8 still use 255 elements, thereby losing any area benefit. Figure 7.9 uses only eight elements. Obviously this converter needs only 1/32 of the area. If every individual element is still of the same size as in the previous examples, it will show a standard deviation of 1 % of its nominal value. In order to design a digital-to-analog converter the eight elements are scaled up according to an exponential series: 1, 2, 4, 8, 16, 32, 64, and 128. In absolute

numbers this means that the MSB element has a standard deviation of $128\sigma_{A_{LSB}}$, while in the example of Fig. 7.8 the 128 elements forming the MSB have a standard deviation of $\sqrt{128}\sigma_{A_{LSB}}$. The DNL is

$$\sigma_{\text{DNL MSB}} = (2^N - 1)\frac{\sigma_{A_{LSB}}}{A_{LSB}}$$

And its magnitude determines the INL. The result in Fig. 7.9 is obvious: the transition at code 128 shows a DNL $= -1$, and constitutes a missing code.

Binary architectures can be designed with low area and power consumption. But the problems with the transitions limit the performance. Another inherent potential issue with binary coded circuits arises from the non-linear relation between the number of switched blocks and the output (e.g., for "15" four units are switched on, for "16" only one unit, for "17" and "18" two units). Switching errors have no correlation with the code. Dynamic errors may occur as a result.

7.1.7 Segmented: INL and DNL

Both techniques, unary and binary coding, are applied in practical design. In the case of converters with a high resolution the choice seems to be between the large number of units involved in a unary design versus the DNL errors associated with a binary coded digital-to-analog converter. Segmentation is generally applied to circumvent these problems: a converter of N-bit resolution is subdivided into a cascade of two sub-converters of M bits in a unary architecture and $N - M$ bits for the binary supplement. Also partitioning in more than two segments is possible.

The full binary architecture of Fig. 7.8 was changed into a segmented architecture, using seven unary elements to form a basic unary architecture in Fig. 7.10. An additional binary architecture completes this structure to form an 8-bit converter. The DNL errors around the unary steps are much less. The main error occurs at the seven codes 32, 64, 96, 128, 160, 192, and 224 where the group of 31 binary elements switches to the 32 unary elements. The DNL will occur at the switching from the full binary structure to the next unary element. For a single transition:

$$\sigma_{\text{DNL}} = \sqrt{2^{N_{binary}+1} - 1}\frac{\sigma_{A_{LSB}}}{A_{LSB}}$$

The resolution of a segmented digital to analog converter is now $N = N_{unary} + N_{binary}$, where N_{unary} is the resolution of the unary section and N_{binary} of the binary section. The overall DNL is determined by the combination of $(2^{N_{unary}} - 1)$ occurrences of the single transition DNL. If the yield for the total converter is required to reach $Y = 1 - Y_{loss}$, where the yield loss $Y_{loss} \ll 1$, then the yield per unary–binary transition must exceed $1 - Y_{loss}/2^{N_{unary}}$. This yield corresponds for a Gaussian distribution with $\mu = 0, \sigma = 1$ to the area contained between the double

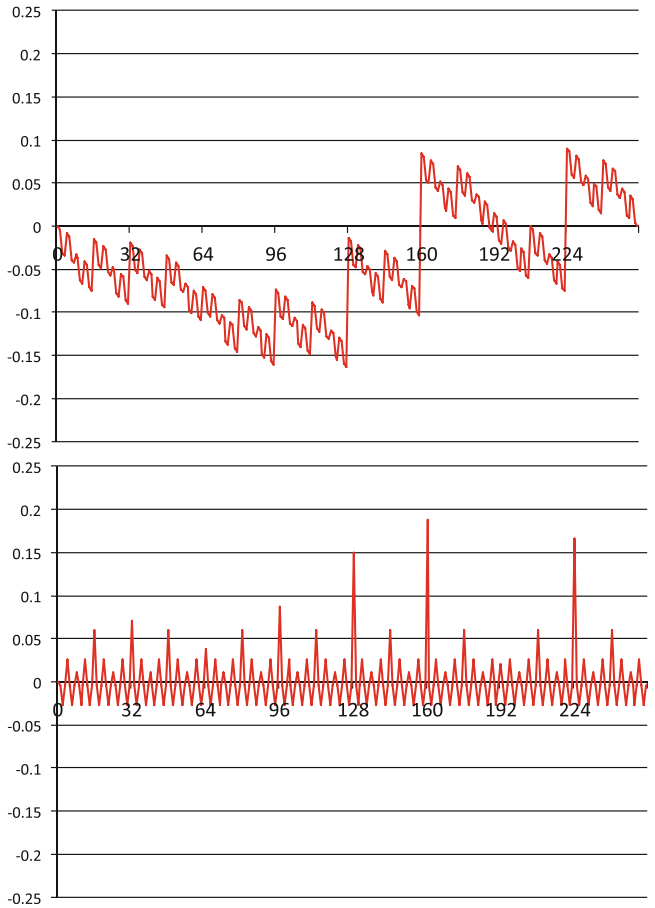


Fig. 7.10 A total of 255 elements are grouped to form a segmented architecture. Seven groups of 32 elements form the MSB unary part and 31 elements are wired to form a binary converter. The integral nonlinearity curve of this simulated 8-bit segmented digital-to-analog converter shows the typical jumps at the MSB transition, the MSB-1 transition, etc. $N_{unary} = 3$, $N_{binary} = 5$

sided rejection limits in Table 2: $-z\sigma, \ldots, +z\sigma$. A yield loss per transition of 0.6 parts-per-million requires $z = 5$ and a performance request of $|\text{DNL}| < 0.5$ leads to $\sigma_{\text{DNL}} = 0.1$ LSB.

The critical elements for the INL variance are the unary elements with value A_{unary} and standard deviation $\sigma_{A_{unary}}$. The effect of this variance on the INL is estimated with the help of Eq. 7.11. However, as the binary section increases the resolution, the result of Eq. 7.11 must be scaled with $2^{N_{binary}}$. Now the standard deviation in the middle is

$$\sigma_{\text{INL}} = \left(2^{N_{unary}/2} 2^{N_{binary}} 2^{-1}\right) \frac{\sigma_{A_{unary}}}{A_{unary}} \tag{7.16}$$

In case the unary elements A_{unary} are constructed of $2^{N_{binary}}$ LSB elements, the substitution of $A_{unary} = 2^{N_{binary}} A_{LSB}$ and $\sigma^2_{A_{unary}} = 2^{N_{binary}} \sigma^2_{A_{LSB}}$ leads to Eq. 7.11.

The observation in Fig. 7.10 that INL in all eight binary sections tends to increase is due to the fact that the random generator in the simulation has set the sum of the binary elements somewhat too large for the average unary element.

Examples of segmentation are found in Sects. 7.3.1 and 7.2.7. Many implementations have been reported in the older literature [130–135], illustrating these basic concepts.

Most digital-to-analog (sub)schemes can be classified along the above architectural split. There are a few deviating forms such as ternary coding (+1,0,-1) that are sometimes used in combination with sign/magnitude representation.

From Figs. 7.5, 7.8, and 7.10 it becomes clear that in either architecture the INL is mainly determined by the area spent. The DNL is, on the other hand, a function of the unary–binary split. The increased DNL for binary architectures is balanced against the individual wiring for each element in a unary architecture.

7.1.8 Architectures in the Physical Domain

In the physical domain the output value of a digital-to-analog converter can be formed using voltages, currents, charges, or time. In each of these physical or analog domains both unary and binary architectures can be used, see Table 7.3.

Voltages can be subdivided by means of resistors. The upper left scheme of Fig. 7.11 shows the concept: a digital decoder selects one of the switches that determines the output voltage. Unary architectures are applied as resistor strings and serve as $N \leq 10$ bit digital-to-analog converters or in flash analog-to-digital converters. $R - 2R$ conversion is mostly applied in low-performance and low-area conversion.

Table 7.3 Various forms of analog representation and physical domains

	Unary	Binary
Voltage	Resistor string *Flash ADC*	R-2R *Low-performance DAC*
Current	Current matrix	Current splitting
	High bandwidth DAC	
Charge/capacitor	Capacitor bank *Low power DAC*	Capacitor bank
Time	PWM, $\Sigma\Delta$ mod *Low bandwidth DAC*	Limited by distortion

In italic the main application area is indicated

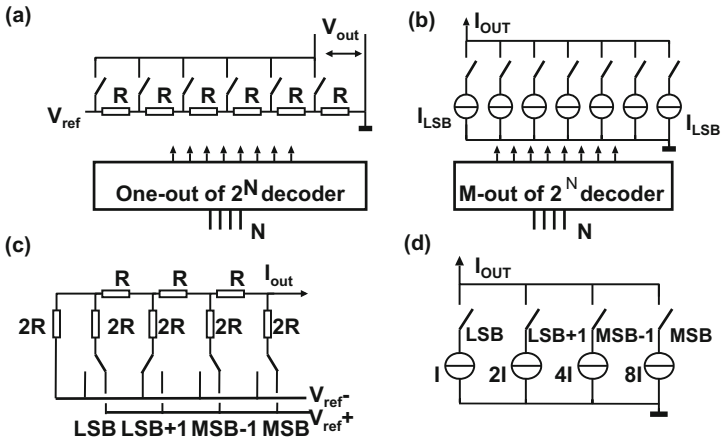


Fig. 7.11 Unary and binary forms of resistor string and current source digital-to-analog conversion. Unary resistor (*upper left*) and unary current source (*upper right*) topologies. Binary resistor (*lower left*) and binary current source (*lower right*) based topologies

In a similar manner as a resistor string, a row of current sources and switches implement a unary current source digital-to-analog converter. Currents simply sum by connecting them together, so a unary row of switched currents is easily complemented with a few binary weighted sources to implement the LSBs. This current-steering architecture is an example of a segmented digital-to-analog converter and is the main implementation form of stand-alone high-performance digital-to-analog converters.

Converters operating in the charge domain use capacitor banks and unary implementations in the time domain use pulse trains, that switch on or off a physical unit.

Capacitor arrays can use the R-2R principle but are limited due to parasitic capacitors.

Using pulses of exponentially increasing lengths in the time domain is realizable; however, it is unclear what advantage that brings.

Next to the combination of signal representation and physical domains Table 7.3 shows the major application area. Except for binary weighted timing all principles find usage.

7.2 Digital-to-Analog Conversion in the Voltage Domain

7.2.1 Resistor Strings

The most used unary digital-to-analog converter in the voltage domain consists of a series connection of resistors between two reference voltages, see Fig. 7.12. These

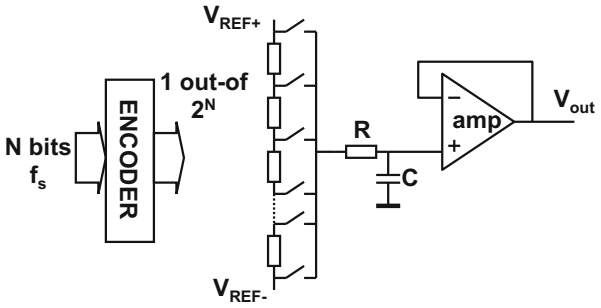


Fig. 7.12 A digital-to analog converter based on a resistive divider

structures are called resistor ladders or resistor strings. A resistor ladder can be turned into a digital-to-analog converter if a selection and switching network is connected to the resistor ladder taps. A buffer is needed to connect the selected tap to the load with a sufficiently low output impedance. A resistor ladder can also serve to generate the quantization levels for a "flash" analog-to-digital converter as discussed in Sect. 8.2.

An important problem in this structure is the variation of the impedance of the resistor ladder: on both ends the equivalent impedance $R_{eq}(m)$ is equal to the impedance of the reference source and close to zero. At a code close to the middle of the ladder the impedance is equal to the parallel connection of two half-ladders. If $m = 0 \ldots 2^N$ is the position of the nodes in a ladder with 2^N resistors,[4] the impedance on each node is

$$R_{eq}(m) = \frac{\frac{m}{2^N}R_{tot} \times \frac{2^N - m}{2^N}R_{tot}}{\frac{m}{2^N}R_{tot} + \frac{2^N - m}{2^N}R_{tot}} = \frac{m(2^N - m)}{2^{N+1}}R_{tot} \tag{7.17}$$

Figure 7.13 shows the parabolic behavior of the effective impedance as a function of the position, with a maximum in the middle. The current that the ladder impedance can deliver will be position and signal-value dependent. With a fixed capacitive load the time constant is position dependent and high-frequency signals will show distortion.

The absolute value and the variation on the ladder impedance require for most applications buffering of the output of the resistor string voltage. The remaining time constant of the resistor string and the capacitance at the input of the buffer must be kept as low as possible to reduce the code-dependent distortion.

[4]When counting elements and nodes of a series string, there is ambiguity: a string of 2^N elements has $2^N + 1$ nodes if the outer connecting nodes are counted. So 1024 resistors give 1025 node voltages. In this book the number of elements is a power of 2, and the highest tap is left unused.

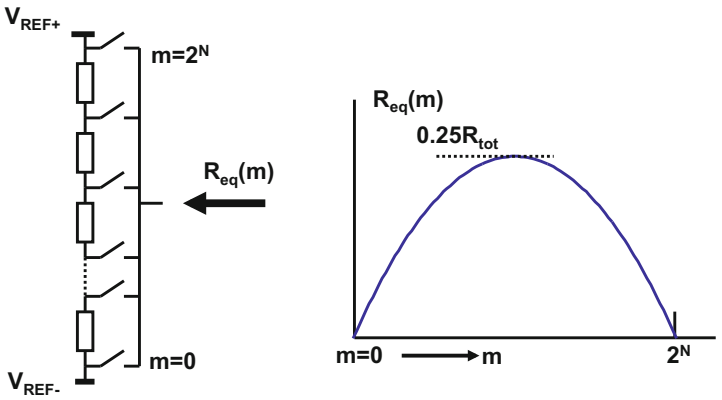


Fig. 7.13 The impedance of a resistor string various with the position

7.2.2 Dynamic Behavior of the Resistor Ladder

The resistor string itself will show capacitive loading, due to parasitical capacitances. This capacitive loading is distributed over the ladder. Excitations by spurious voltages or charges will cause a settling behavior. In order to facilitate an analysis, the resistor string is modeled as a continuous resistive strip, where the resistance and capacitance per unit length are defined by r in Ω/m and c in F/m. The voltage over this strip with length L is described by the diffusion equation. This equation is also referred to as "heat equation"[5] and describes in classical thermodynamic theory the evolution of the temperature as a function of time and position:

$$\frac{\partial \text{Temperature}(x,t)}{\partial t} = D\frac{\partial^2 \text{Temperature}(x,t)}{\partial x^2} \tag{7.18}$$

where D is the thermal diffusion constant. This equation is used in the voltage domain with the function $v(x,t)$ describing the voltage in time and position over the resistive structure:

$$rc\frac{\partial v(x,t)}{\partial t} = \frac{\partial^2 v(x,t)}{\partial x^2} \tag{7.19}$$

With the boundary conditions at $v(x,0) = v_{start}(x)$ and at $v(0,t) = v(L,t) = 0$, an exact solution can be obtained of the form:

[5]It is convenient to look in literature for solutions of the "Heat equation" problem with your specific boundary conditions and rewrite them to voltage equations.

$$v(x,t) = \sum_{k=1}^{\infty} e^{-k^2\pi^2 t/rcL^2} a_k \sin(\frac{k\pi x}{L}) \tag{7.20}$$

The solution is orthogonal for time and position, both are described by separate functions. The start condition is brought into the equation by solving the equation for $t = 0$:

$$v(x,0) = v_{start}(x) = \sum_{k=1}^{\infty} a_k \sin(\frac{k\pi x}{L}) \tag{7.21}$$

The terms a_k with the sin function are a Fourier description in one dimension.

The initial condition that is defined by $v_{start}(x)$ will exponentially decay. The decay behavior is dominated by the first term ($k = 1$):

$$v(x_m,t) \approx v_{start}(x_m) e^{-\pi^2 t/rcL^2} \tag{7.22}$$

with a time constant: $\tau = rcL^2/\pi^2$. Note that $R_{tot} = rL$, $C_{tot} = cL$.

Looking at this problem from an engineering point of view, the starting observation is that the largest errors due to the distributed delay line will occur in the middle of the structure. From this point the resistive impedance towards the ends is $R_{tot}/2$ parallel to $R_{tot}/2$, or $R_{tot}/4 = rL/4$. The capacitors at the intermediate nodes need to be charged. The capacitors close to the middle tap must be fully charged, and capacitors closer to the ends will be charged proportional to their position. On average the total capacitance is charged to half the value of the middle tap. That totals to a time constant on the middle tap of: $\tau = R_{tot}C_{tot}/8 = rcL^2/8$, which is not equal to the solution of the heat equation, but close enough for a first order guess. Figure 7.14 compares the approximation with the distributed line solution.

7.2.3 Practical Issues in Resistor Ladders

The actual design in the lay-out of a resistor ladder requires to consider various aspects. The available resistive material must show low random variation, a zero-voltage and zero-temperature coefficients and a low parasitic capacitance. The settling time constant $\tau = R_{tot}C_{tot}/\pi^2$ must be low enough for the required conversion time. For example, for high-speed conversion >100 Ms/s and 10-bit resolution, the time constant must be $\tau < 1$ ns. As the total parasitic capacitance is significant, extremely low values for the tap resistances are necessary. Realizing tap resistances in the order of magnitude of 1 Ω poses considerable (lay-out) problems and requires sometimes the use of special material layers on top of the chip.

See Sects. 5.2.3 and 5.3 for some more background on technological effects.

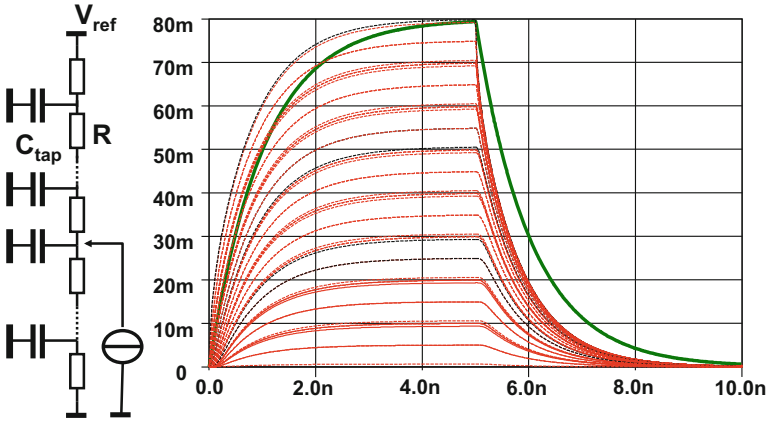


Fig. 7.14 In a ladder string with 256 resistors of 1.25 Ω and $C_{tap} = 0.1$ pF each, a current is injected at the middle tap at $t = 0$ and switched off after 5 ns. The *thin-line plots* show the time behavior at various nodes in the ladder with a time constant: $\tau = R_{tot}C_{tot}/\pi^2 = 320\,\Omega \times 25.6\,\text{pF}/\pi^2 = 0.83$ ns. The *bold line* is the lumped RC approximation with $\tau = R_{tot}C_{tot}/8 = 1.024$ ns

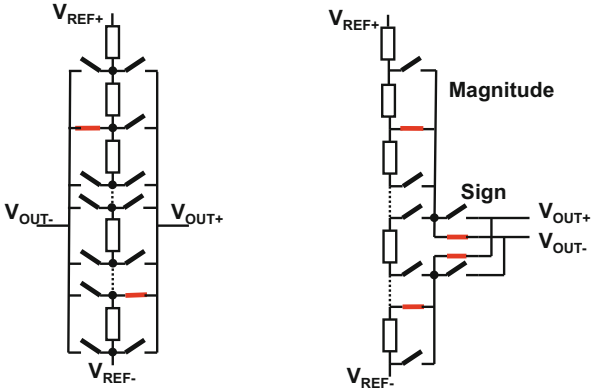


Fig. 7.15 A differential ladder structure and a sign-and-magnitude topology

An alternative way to eliminate a first order gradient is to use a differential signal. Figure 7.15 shows two examples: a ladder structure with full differential decoding and a structure based on a sign-and-magnitude decoder.

Resistor ladders are crucial as building blocks in flash analog-to-digital converters. When a resistor ladder converter is applied as an output device for driving an application, a careful design of the buffer is required, see Sect. 7.2.7. This topology allows an excellent static performance at a limited power level. The speed is limited by the performance of the buffer.

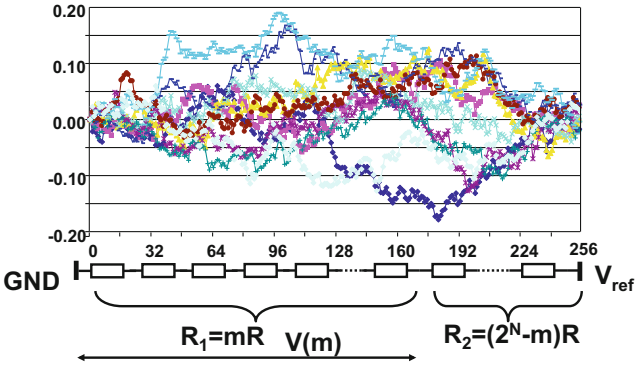


Fig. 7.16 A resistor string is connected between a reference voltage and ground. The simulation shows ten instances of the effect of $\sigma_R/R = 1\,\%$ mismatch. In this 256 devices ladder the maximum deviation is $\pm 20\,\%$

7.2.4 Accuracy in Resistors Strings

The accuracy in which a quantity can be subdivided limits the obtainable performance of a digital-to-analog converter. In Fig. 7.16 a ladder of 256 equally designed resistors is connected between a reference voltage and ground. Ideally the voltage at the m-th position $m = 0, 1, \ldots, 2^N$ is

$$V(m) = \frac{m}{2^N} V_{ref} = \frac{mR}{mR + (2^N - m)R} V_{ref} = \frac{R_1}{R_1 + R_2} V_{ref} \tag{7.23}$$

where R_1 and R_2 represent the resistance up to position m and the remaining resistance. Although devices can be equally designed, some random variation is unavoidable. If all resistor values are subject to a mutually independent random process with normal distributions with a mean value R and a variance σ_R^2, the variance in the voltage at the m-th position can be calculated.

The resistance R_1 is a string of m resistors and shows an expectation value and variance:

$$E(R_1(m)) = E(mR) = mE(R) = mR, \quad \sigma_{R1}^2 = m\sigma_R^2 \tag{7.24}$$

The other quantity involved R_2 is the string of $(2^N - m)$ resistors which is independent of the complementary string R_1. For the string R_2 the mean and the variance are found in a similar way.[6] The variance of the voltage $V(m)$ is now found by applying the statistics formula for multiple stochastic variables, Eq. 5.11:

[6]It seems that a shortcut is possible by using the string of M resistors, however this string shares m resistors with R_1 and the covariance has to be included, which is a possible route, but not pleasant.

$$\begin{aligned}\sigma_V^2(m) &= \left(\frac{\partial V(m)}{\partial R_1}\right)^2 \sigma_{R1}^2 + \left(\frac{\partial V(m)}{\partial R_2}\right)^2 \sigma_{R2}^2 \\ &= \left(\frac{R_2}{(R_1+R_2)^2}\right)^2 \sigma_{R1}^2 V_{ref}^2 + \left(\frac{-R_1}{(R_1+R_2)^2}\right)^2 \sigma_{R2}^2 V_{ref}^2 \\ &= \frac{m(2^N-m)}{2^N}\frac{\sigma_R^2}{R^2}\frac{V_{ref}^2}{2^{2N}} = \frac{m(2^N-m)}{2^N}\frac{\sigma_R^2}{R^2}V_{LSB}^2\end{aligned}$$

with $V_{ref} = 2^N \times V_{LSB}$. Compare also to Eq. 7.17 where the ladder impedance is calculated: the impedance dependence on the position is mathematically similar to the random variance. The maximum value of the variance occurs at $m = 2^{N-1}$ for which the variance is found as

$$\sigma_V^2(m = 2^{N-1}) = \frac{1}{2^{N+2}}\frac{\sigma_R^2}{R^2}V_{ref}^2 = 2^{N-2}\frac{\sigma_R^2}{R^2}V_{LSB}^2 \tag{7.25}$$

The statistical component in the INL at position $m = 0, 1, \ldots, 2^N$ is $\sigma_V(m)$ normalized to V_{LSB}:

$$\sigma_{INL}^2(m) = \frac{m(2^N-m)}{2^N}\frac{\sigma_R^2}{R^2} \tag{7.26}$$

with a maximum in the middle at $m = 2^{N-1}$ of:

$$\sigma_{INL}(2^{N-1}) = \sqrt{2^{N-1}}\frac{\sigma_R}{R} \tag{7.27}$$

The position of the maximum value of the deviation in an arbitrary ladder structure is not necessarily in the middle of the string, although the probability is there the highest. Nevertheless Eq. 7.25 can serve to estimate the INL. With the help of Table 2 the sigma margin can be chosen.

In the example of Fig. 7.16 the ladder contains 256 resistors with a relative resistor mismatch of 1 %. The relative sigma value in the middle of this ladder is therefore 8 % and in the 10 random simulations, excursions up to 20 % of an LSB are seen. These values directly impact the integral linearity.

The differential linearity is given by the variation in the step size itself and equal to the expected maximum deviation of one resistor. The DNL of a resistor string is determined by the single resistor variance and the number of resistors. The maximum DNL is determined by the largest deviating resistors amongst 2^N random varying resistors with each a relative spread of σ_R/R. In order to guarantee the DNL for a production lot the estimation will have to be extended to P converters times 2^N resistors per converter.

Example 7.1. If the resistors in a string show a relative spread of 1 %, what is the probability that one resistor in a string of 256 exceeds a deviation of 4 %?

Solution. The probability p that one resistor with value R_a and $\sigma_R/R = 1\,\%$ deviates more than 4 % from the average value R is

$$p = P\left(\left|\frac{R_a - R}{R}\right| > \frac{4\sigma_R}{R}\right) = P(|x| > 4\sigma) \tag{7.28}$$

with Table 2, $p = 6.3\times10^{-5}$. The yield per resistor is $(1-p)$. The yield for a resistor string with $2^N = 256$ resistors is $(1-p)^{256} = 98.4\,\%$.

Example 7.2. 65 current sources of 0.1 mA each are arranged in a line and connected with their negative terminal to an aluminum return wire. This wire shows a 0.1 Ω impedance between two adjacent current sources. So the total line impedance is 6.4 Ω. Calculate the maximum voltage drop over this return line, if:

- the return line is connected to ground at one extreme,
- the return line is connected to ground at both extremes,
- the return line is connected to ground in the middle.

Solution. This problem bears similarity to the calculation of the time constant on the distributed *RC* ladder. Starting at the open end of the return wire, it is clear that the current source causes a voltage drop *IR* over the first segment of the return wire. The voltage drop over the second segment is double, over the third segment triple, etc. The total voltage drop is therefore: $(1+2+3+\ldots+n)IR$. The sum of this series is: $IRn(n+1)/2$. With the data above the voltage drop at the open end is 20.8 mV.

If the open end is also grounded, the maximum voltage will appear in the middle. The middle current source contributes half to both sides so the voltage drop is: $32\times0.5IR + \sum_{i=1}^{i=31} IR = 5.12\,\text{mV}$.

One ground connection in the middle exactly splits the first problem into two, so at both open ends a voltage appears of $\sum_{i=1}^{i=32} IR = 5.28\,\text{mV}$.

Example 7.3. Show that a perfect linear gradient in Fig. 7.15 is canceled by the differential read-out of a single ladder.

Solution. If $m = 0,\ldots,(2^N-1)$ is the digital code, then a perfect ladder will produce a voltage on every resistor node: $V(m) = V_{REF-} + mV_{LSB}$, where $V_{LSB} = 2^{-N}(V_{REF+} - V_{REF-})$. A linear gradient means that in the direction of increasing resistivity every next LSB is slightly (ΔV_{LSB}) larger. Now the voltage on every node is: $V(m) = V_{REF-} + m(V_{LSB} + m\Delta V_{LSB})/S$. The term between brackets must be scaled back with a factor $S = (1 + 2^N\Delta V_{LSB}/V_{LSB})$ to fit to the reference voltages.

Now a differential output voltage can be formed by choosing the node voltage connected to m and the complementary code connected at tap $2^N - m$. This gives as a differential output voltage:

$$\begin{aligned} &V_{OUT+}(m) - V_{OUT-}(2^N - m) = \\ &V_{REF-} + m(V_{LSB} + m\Delta V_{LSB})/S - \left[V_{REF-} + (2^N - m)(V_{LSB} + (2^N - m)\Delta V_{LSB})/S\right] \\ &\quad = m(2V_{LSB} + 2^{N+1}\Delta V_{LSB}) - (2^N V_{LSB} - 2^{2N}\Delta V_{LSB})/S \end{aligned}$$

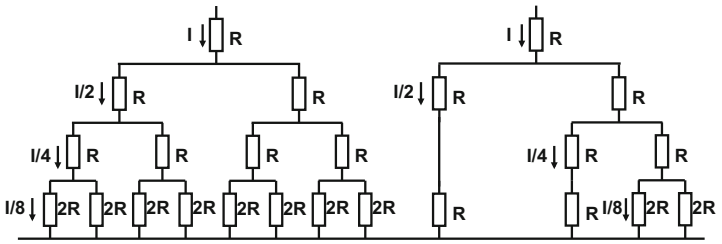


Fig. 7.17 Currents can be split in binary weighted portions (*left*). The constant impedance in each branch allows to reduce the branches of the structure

There are only linear signal components (proportional to i) left in the input signal. Note that the common mode signal $(V_{OUT+} + V_{OUT-})/2$ contains the second order distortion term. So the common mode rejection in the succeeding stages is important to prevent that the distortion in the common mode signal enters the differential path.

7.2.5 R-2R Ladders

Unary resistor structures require 2^N elements to implement an N-bit converter. The size of such a structure doubles for every bit and the settling time increases fourfold. Figure 7.17 shows on the left side a binary conversion structure built with resistors of value R. This tree is terminated with resistors of value $2R$. In every layer twice the number of currents flow, each of half of the value of the layer above. Moreover from each node, looking downwards, the impedance equals R. This property allows to replace in Fig. 7.17 (right) the branches by resistors R and thereby constructing the $R-2R$ structure.[7] Figure 7.18 shows a more abstract view and a realization of a binary coded digital-to-analog converter based on the "R-2R" principle. Current entering the R-2R resistor circuit splits at every node in two equal parts, because the impedance in both directions equals 2R. The combination of branches therefore generates a power-of-two series of currents, that can be combined via switches into an output current. In Fig. 7.18 this current is fed into an opamp fed back with a resistor. The unused current has to be dumped into ground.

The simple analysis of an $R-2R$ ladder starts from the LSB side. The resistor network is terminated with a resistance $2R$. This resistance is put in parallel to an equally sized value that generates the LSB current. The combined value of the two parallel resistors is R. If the resistor in series is added, the total impedance of the section is back to the original $2R$. This impedance is connected in parallel with the $2R$ resistor for the LSB-1 section and again yields an R value.

[7] Explanation from Colin Lyden (ADI).

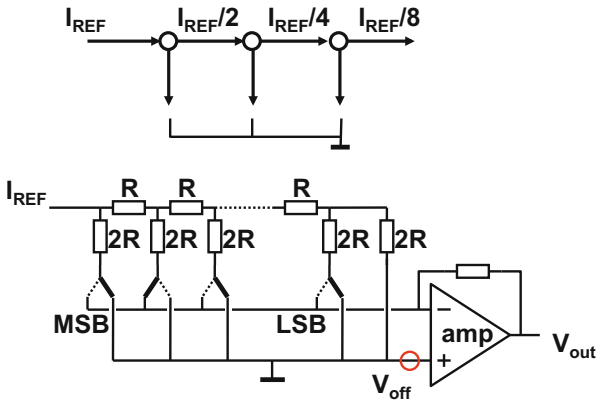


Fig. 7.18 R-2R digital-to-analog converter based on the $R-2R$ principle with very simple digital encoding. Offset voltages of the virtual ground and switch impedances limit the accuracy

A buffer is used to convert the current from the R-2R ladder into a suited output format. The bandwidth of the buffer is limiting the overall bandwidth of this converter. This technique allows to design digital-to-analog converters of a reasonable quality (8–10 bits) at low power consumption and low area. Moreover the digital coding can be directly taken from a straight binary representation. The application of simple $R-2R$ converters is limited to low-resolution low-cost applications such as offset correction.

7.2.6 Accuracy in R-2R Structures

The main accuracy problem in an R-2R digital-to-analog converter is caused by the inequality in the splitting process. The main reasons are voltage offsets between the ground line and the summation node, resistor inequality and the switch.

A voltage difference between the ground for the redundant resistor currents and the virtual ground of the opamp will affect the current splitting. If this offset is denoted as V_{off}, the current deviation equals $V_{off}/2R$ when the MSB is active. If all bits are active except for the MSB, the error has the opposite sign. Consequently the offset must be smaller than $V_{off} < I_{REF}R2^{-N}$.

The resistors themselves must match to the required accuracy level, in contrast to the unary resistor string. If the impedance in the first vertical branch has a 1 % larger resistance, the current will split in 0.495 and 0.505 portions, limiting the achievable resolution (for a $|DNL| < 1$) to 7 bit.

The resistor structure needs various connections through which current is flowing. Technological contacts are a major source for variability.

The main reason for inequality between the branches are the switches. As the two voltages to which the resistors are switched vastly differ, the switch to ground

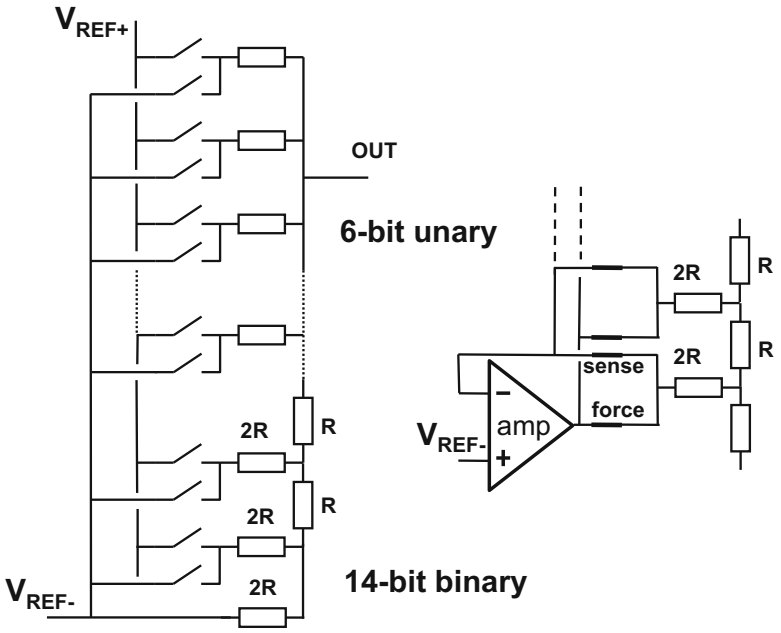


Fig. 7.19 A combination of unary and R-2R resistors allows to build a digital-to-analog with 20-bit DC accuracy [138]

is preferably implemented as an NMOS transistor, while the switch to the positive reference voltage is a PMOS device. The transistors are scaled according to the binary weight of the branch and the impedance of the switch can be part of the $2R$ resistor. Still that is not optimum for matching as the drive voltages and the type of transistors differ. One way out is to increase the unit resistor to the 100 kΩ range, which makes the switch impedance negligible [136].

In the implementation scheme of Sect. 7.2.7 the switch impedance is made equal by means of an additional voltage divider, see Fig. 7.20. Others [137] have proposed ideas to measure the impedances and keep them equal via a control loop.

A rigorous solution to the problem is presented in [138], see Fig. 7.19. The authors apply thin-film resistors with excellent properties. The segmented architecture consists of a 6-bit unary digital-to-analog converter combined with a 14-bit R-2R structure. The converter is driven by ±10 V. In order to mitigate the problem with the switch resistances, a sense-force topology is used.[8] Figure 7.19 (right) shows an opamp in a loop: in the sense feedback path there is no current, so the opamp will force the sensed node to the reference voltage. A similar structure is used for the positive side. Any opamp offset shows up as reference voltage error, which is easily correctable. After calibration an INL performance of 20-bit is reported.

[8] Also known as Kelvin-connection and four-point sensing.

7.2.7 *Implementation of a Video Resistor-Ladder Digital-to-Analog Converter*

In a system chip with analog-to-digital and digital-to-analog converters, it is advantageous to have similar references for the analog-to-digital and digital-to-analog converters. The tracking of input and output ranges for processing variations, temperature, etc. is then guaranteed and the overall gain of analog-to-digital and digital-to-analog converters is better controlled. The voltage dependence and the mutual matching of large-area polysilicon resistors[9] allow the design of a converter with high integral and differential linearity.

The design of a digital-to-analog converter with a single 1024-tap resistor ladder and sufficiently fast output settling requires a tap resistance in the order of 6–10 Ω. The size of such resistors in conventional polysilicon technology is such that accurate resistor matching, and hence linearity, becomes a problem. The applied solution in this design is to use a dual ladder [139] with a matrix organization [140]. Figure 7.20 shows the ladder structure: the coarse ladder consists of two ladders, each with 2 large-area resistors of 250 Ω with 16 taps, which are connected anti-parallel to eliminate the first-order resistivity and temperature gradients. The coarse ladder determines 16 accurate tap voltages and is responsible for the integral

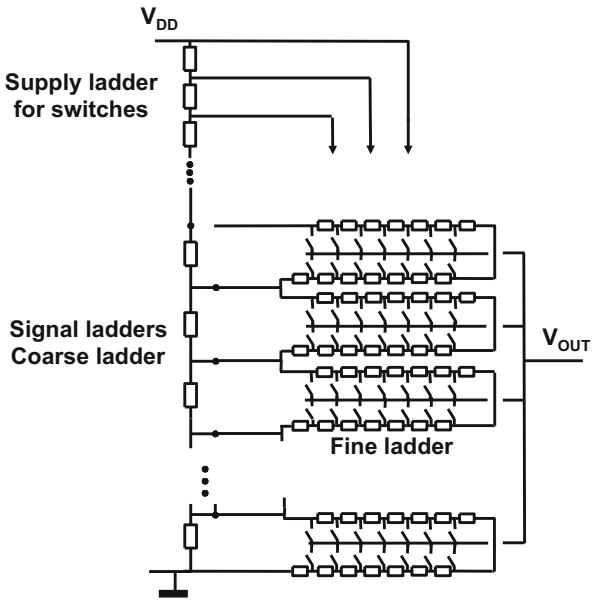


Fig. 7.20 Resistor network for a video digital-to-analog converter [141]

[9] Diffused resistors are a preferred alternative in more advanced processes.

linearity. A 1024-resistor fine ladder is arranged in a 32-by-32 matrix, in which every 64th tap is connected to the coarse ladder taps (see Fig. 7.14). The fine ladder tap resistance is chosen at 75 Ω without loss of speed. The wiring resistances can be neglected compared to the 75 Ω tap resistors. There are only currents in the connections between the ladders in the case of ladder inequalities; this reduces the effect of contact resistance variance. The current density in the polysilicon is kept constant to avoid field-dependent non-linearities. The coarse ladder is designed with polysilicon resistors in order to avoid voltage dependence of diffused resistors. The fine ladder is designed in either polysilicon or diffusion, depending on secondary effects in the process implementation.

The second source of the varying output impedance is the switch transistor; usually its gate voltage equals the positive power supply, but the voltage on its source terminal is position dependent. The turn-on voltage doubles from one end of the ladder to the other. In this design an additional supply ladder is placed on top of the signal ladders to keep the turn-on voltage of the switches more constant. The turn-on voltage of each switch transistor is effectively made to correspond to the lowest turn-on voltage of the ladder digital-to-analog structure.

The total ladder configuration can still be fed from the 3.3 V analog power supply; the signal ladders are in the range between ground level and $0.4V_{DD}$, the supply ladder goes from $0.6V_{DD}$ to V_{DD}.

The core of the digital-to-analog converter is formed by the 32-by-32 fine-resistor matrix. The two decoders are placed on two sides of the matrix. The two sets of 32 decoded lines are latched by the main clock before running horizontally and vertically over the matrix. In the matrix, the 1024 AND gates perform the final decoding from the 32 horizontal MSB lines and the 32 vertical LSB lines. The insert in Fig. 7.21 shows a detail of the matrix decoding.

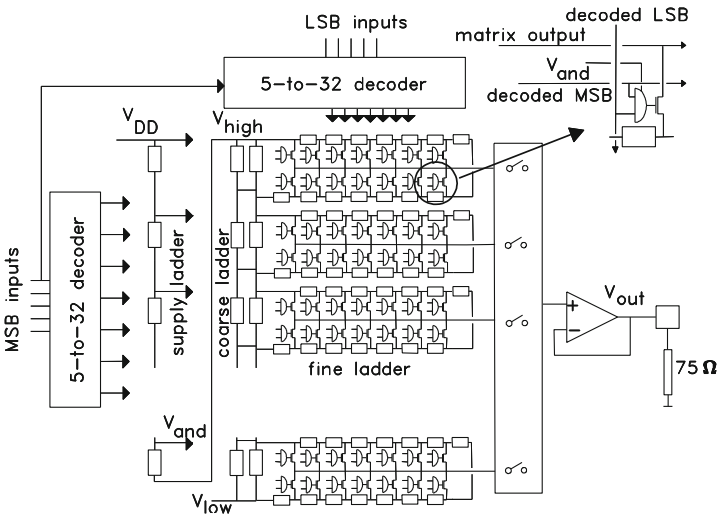


Fig. 7.21 Block diagram of the digital-to-analog converter [141]

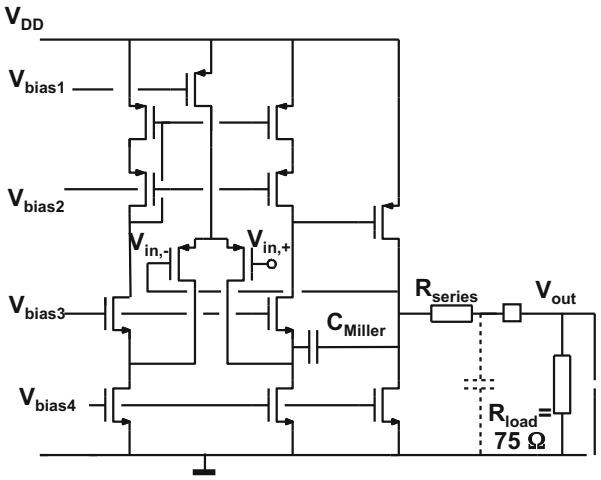


Fig. 7.22 Folded cascode opamp circuit with Miller compensation used for the buffer

A voltage-domain digital-to-analog converter generates only the necessary momentary current and is hence more efficient in power consumption. However, a voltage-domain digital-to-analog converter requires an on-chip buffer, which introduces two drawbacks: the output always needs some offset from the power supply rail and the opamp is inherently slower due to its internal pole. The output buffer is a folded-cascode opamp with Miller compensation. The Miller feedback current flows via the source of the cascode transistor to the high-ohmic node [142]. This approach avoids a stop resistor in series with the Miller capacitor. The stop resistor is not desired because the transconductance of PMOS driver is not very constant.

The p-channel input stage operates on input voltages ranging from 0 to 2.2 V, see Fig. 7.22. The main current path through the output stage goes from the PMOS driver (W/L = 1400) down into the output load. A resistive load is consequently needed for optimum performance.

The on-chip stop resistor R_{series} is on the order of 25–75 Ω; it keeps up a feedback path even at frequencies at which the bond pad capacitance shorts the circuit output. It also serves as a line termination. The voltage swing over the output load resistor is consequently half of the buffer input voltage. The actual value of the stop-resistor can be controlled to within 10 %; the resulting gain error is no serious drawback in video equipment, as there is always an overall gain adjustment.

Figure 7.23 shows a photograph of a triple 10-bit converter. The power distribution over the 10-bit digital-to-analog converter is dominated by the output stage: with a full-swing sine wave (0.1–1.1 V on 75 Ω), the average current through the driver transistor is 7.3 mA. The remaining part of the driver requires 1 mA. The ladder current is 1 mA while the digital part running at 50 MHz is limited to 0.7 mA, resulting in a total power supply current of 10 mA. The design in a 1 µm technology achieves a DNL of 0.1 LSB.

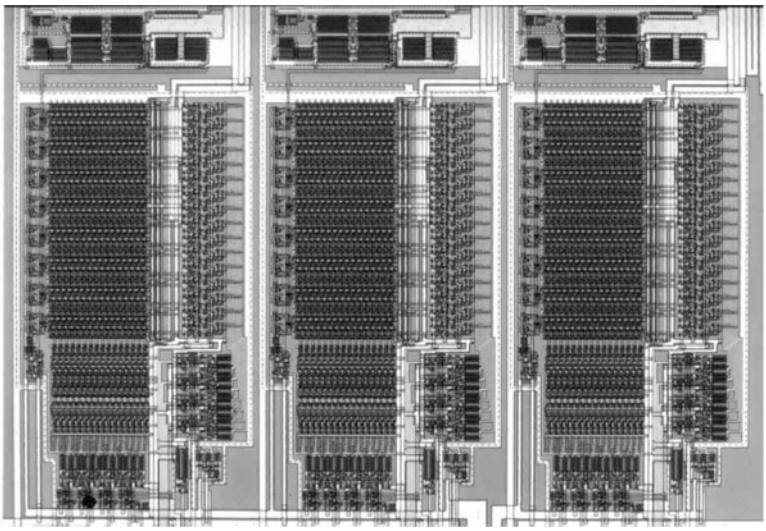


Fig. 7.23 This 10-bit triple digital–analog converter based on voltage dividers is used for applications in the video domain for the basic video colors: red, green, and blue

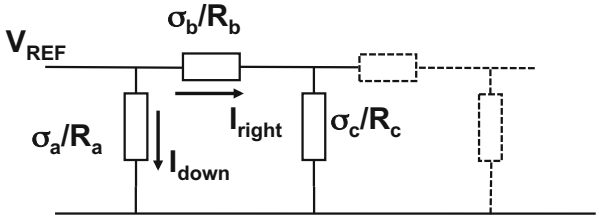


Fig. 7.24 *Left*: basic section of R-2R ladder, *right*: current splitting at the first node

Table 7.4 in Sect. 7.3.6 summarizes the performance [141].

Example 7.4. In an R-2R ladder each resistor has an uncorrelated standard deviation of 1 % of its value. If a DA converter is constructed with these resistors, how many bits of resolution can be designed without running into monotonicity problems?

Solution. Examine a basic section of the R-2R ladder in Fig. 7.24. The variance of $(R_b + R_c)$ is: $\sigma_b^2 + \sigma_c^2$. The equation for the current split of the resistors is $cs = (R_a - R_b - R_c)/(R_a + R_b + R_c)$, where R_c stands for the remainder of the structure:

$$\sigma_{cs}^2 = \frac{4(R_b + R_c)^2}{(R_a + R_b + R_c)^4}\sigma_a^2 + \frac{4R_a^2}{(R_a + R_b + R_c)^4}(\sigma_b^2 + \sigma_c^2)$$

An R-2R ladder is designed with: $R_a = 2R, R_b = R_c = R$, which reduces the above equation to: $\sigma_{cs}^2 = (\sigma_a^2 + \sigma_b^2 + \sigma_c^2)/16R^2$. The observation that replacing R_c in the basic schematic by the same three resistor scheme allows to expand the basic schematic towards an R-2R ladder in an iterative way. Using: $1 + r + r^2 + \ldots = 1/(1-r)$ gives: $\sigma_{cs}^2 = (\sigma_a^2 + \sigma_b^2)/15R^2$. With $\sigma_a = 0.02R$ and $\sigma_b = 0.01R$, $\sigma_{cs} = 0.01/\sqrt{3}$.

The influence of the additional sections on the splitting process of the first stage is marginal due to the division by 16.

This allows to reduce the current splitting problem to the current splitting in the first stage. Figure 7.24 shows two equivalent input resistors, each nominal $2R$. Resistor R_a has a specified standard deviation of $\sigma_{Ra} = 0.02R$ and $R_{right} = 2R$ with $\sigma_{Rright} = R\sqrt{0.01^2 + 0.00577^2} = 0.0115R$.

Analyzing the current splitting of $I_{cs} = I_{right} - I_{down}$:

$$\sigma_{Iright-Idown}^2 = \sigma_{Iright}^2 + \sigma_{Idown}^2 = \left(\frac{dI_{right}}{dR_{right}}\right)^2 \sigma_{Rright}^2 + \left(\frac{dI_{down}}{dR_a}\right)^2 \sigma_{Ra}^2 =$$

$$\left(\frac{d}{dR_{right}}\left(\frac{V_{REF}}{R_{right}}\right)\right)^2 \sigma_{Rright}^2 + \left(\frac{d}{dR_a}\left(\frac{V_{REF}}{R_a}\right)\right)^2 \sigma_{Ra}^2 =$$

$$\left(-\frac{V_{REF}}{R_{right}^2}\right)^2 \sigma_{Rright}^2 + \left(-\frac{V_{REF}}{R_a^2}\right)^2 \sigma_{Ra}^2 =$$

$$\frac{I_{tot}}{4R}\left(\sigma_{Rright}^2 + \sigma_{Ra}^2\right) = \frac{I_{tot}}{4}\sqrt{0.0115^2 + 0.02^2} = 0.0058 I_{tot}$$

One σ equals 0.58 % of the total current. Monotonicity means that the maximum error in $I_{right} - I_{down}$ is not exceeding an LSB: $2^{-N}(I_{right} + I_{down})$. For $N = 6$, an LSB equals 1.6 %. So for a 6-bit R-2R DAC there is a 2.8σ probability or 0.5 % chance that a non-monotonicity error will occur.

Example 7.5. In Fig. 7.25 an alternative two-ladder structure is shown. Now only one fine ladder is connected over the relevant coarse ladder segment. Its current is supplied by two current sources on top and bottom. Comment on the differences with respect to Fig. 7.20.

Solution. Obviously the area of this solution is lower than in the converter of Fig. 7.20. That advantage is balanced by quite some problems. For most samples the fine ladder is connected to a new coarse segment. That means that the fine ladder with its connected parasitic capacitances will have to settle. Another issue is the tuning of the two current sources. First of all they have to be equal: $I_{fine,1} = I_{fine,2}$. Any excess current will flow into the coarse ladder creating offsets. Moreover, the currents must be chosen such that $I_{fine,1} \times R_{fine,tot} = V_{coarse}$, where $R_{fine,tot}$ is the total fine ladder resistance. This last requirement must be met within a fraction of an LSB. The resistances of the additional switches between the coarse and fine sections must be low compared to the ladder impedances.

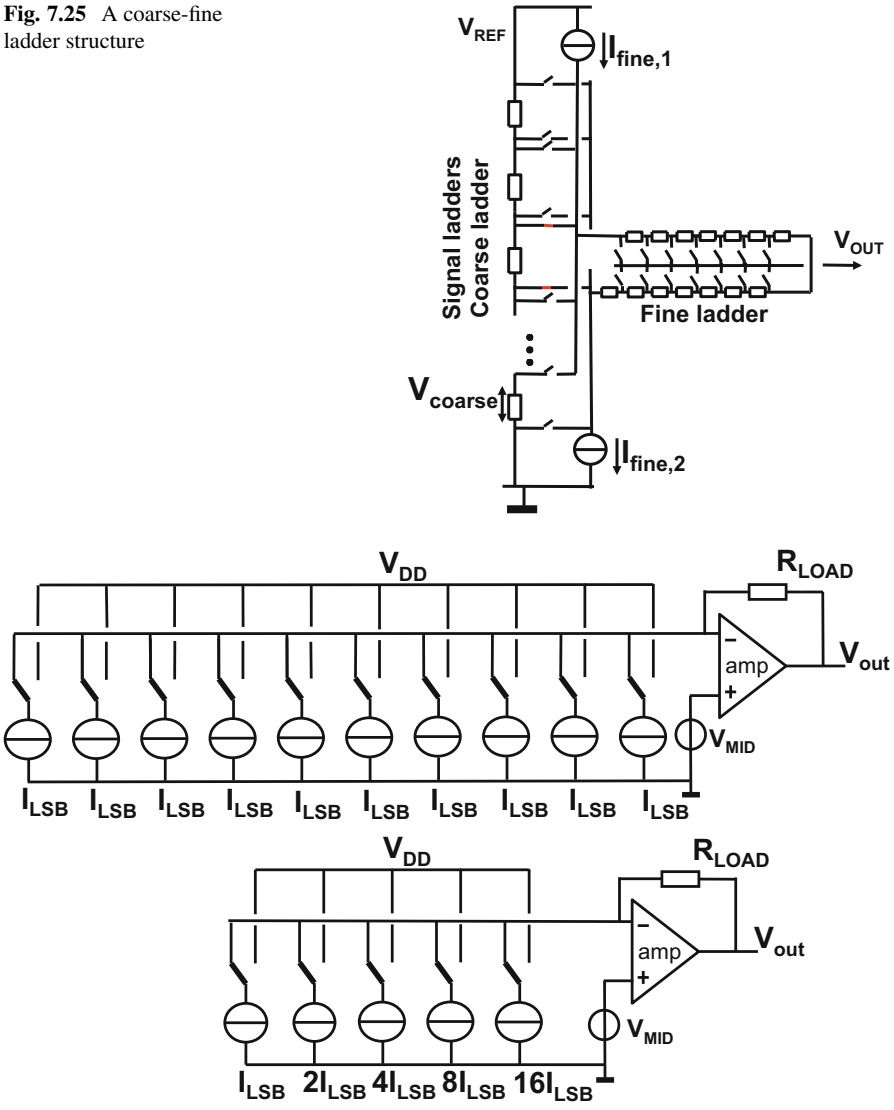


Fig. 7.25 A coarse-fine ladder structure

Fig. 7.26 Digital–analog converter with current sources: unary and binary forms

7.3 Digital-to-Analog Conversion in the Current Domain

7.3.1 Buffered Current Domain Digital-to-Analog Converters

Figure 7.26 shows the block diagrams of unary and binary coded digital-to-analog converters in the current domain with a buffer. The array of current sources can use a unary, binary, or segmented architecture, with the advantages and disadvantages

discussed in Table 7.2. In the unary architecture the current sources are sequentially accessed: if the next digital input word is 5 LSBs larger than the previous, the next five current sources in line are switched to the output rail. In case of the binary architecture the digital word directly controls the switches. An increase in the digital code can lead to switching on or off a few current sources, but also to toggling all switches (e.g. if the code changes from $2^{N-1} - 1$ to 2^{N-1}). The current source requirements are mostly dictated by the required accuracy. As more accuracy requires more transistor area, the current source devices must be designed with large area, see the example on page 171. Unused current sources can be switched off, losing the inversion charge. As it takes time to bring back these current sources into operation, mostly the current sources are kept active and the unused current is dumped in a power rail.

The buffer in Fig. 7.26 provides a low-impedance load for the current sources. The summation line is kept on a fixed voltage avoiding modulation of the currents due to their finite output impedance, which is the dominant problem in the digital-to-analog converters of Sect. 7.3.2. Any load variation is shielded from the current sources. Also there is no strong relation between the level of the current sources and the delivered output current which allows to keep power in the current source array at a modest level.

When symmetrical signals are converted with the structures in Fig. 7.26 a bias level is established corresponding to the zero level of the symmetrical signal. This bias is mostly set at half of the maximum current. So the output has a zero-signal level of $V_{MID} + R_{LOAD} I_{total}/2$, where R_{LOAD} is the feed-back resistor and I_{total} the total current of the array. As current sources generate $1/f$ and thermal noise, the bias level is polluted. Also a zero-signal level will experience a considerable noise level.

Figure 7.27 shows an arrangement with arrays of positive and negative current sources. This architecture has some interesting properties. The structure fits well to the sign-and-magnitude representation and provides a symmetrical signal swing,

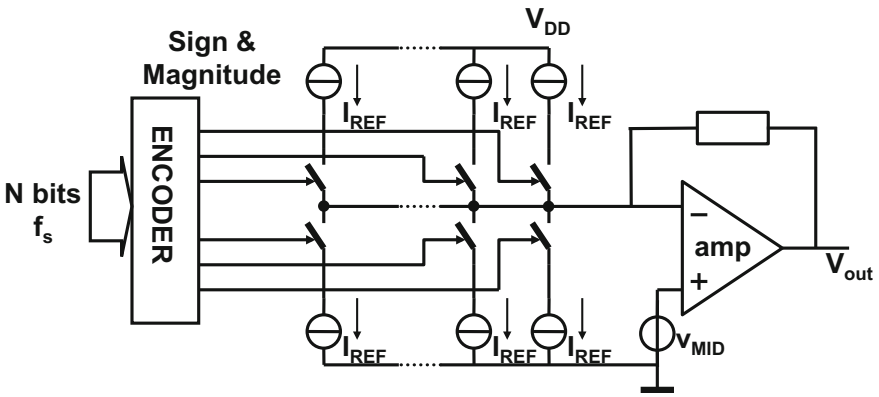


Fig. 7.27 Two-rail digital-to-analog converter with current sources

e.g. a sine wave around a zero-level, in Fig. 7.27 indicated by V_{MID}. For a zero-level signal level the digital-to-analog converter in Fig. 7.27 has no current source connected to the summation line. Only the noise of the buffer will show. This allows an excellent signal-to-noise ratio for small signals and consequently increases the dynamic range of the converter on the low side. For larger signals, the noise contributions increase. Especially in communication protocols, like advanced ADSL schemes, this converter is an interesting implementation choice. This is another example where the specification of the dynamic range (DR) and the signal-to-noise ratio (SNR) differ.

The obvious disadvantage is the risk of inherent inequality between the current sources generating positive signals and the current sources for the negative signals. This issue somehow has to be addressed. Modern calibration techniques are progressing and allow 12-bit accuracy. A full differential implementation is another option as it turns structural inequalities between NMOST and PMOST current sources into an offset and gain error. Major disadvantage of all buffered arrangements (resistors, currents, and capacitors) is the feedback stabilization of the opamp used for the buffer. If the buffer has to drive a load consisting of resistive and capacitive elements, the unity gain of the buffer has to be designed at a frequency lower than the dominant pole of the output load in order to avoid output ringing, Fig. 7.28. Consequently the overall bandwidth is degraded. When the design choice for a buffered output is made, a capacitor array or resistor ladder solution is generally more advantageous as these provide better matching and lower power.

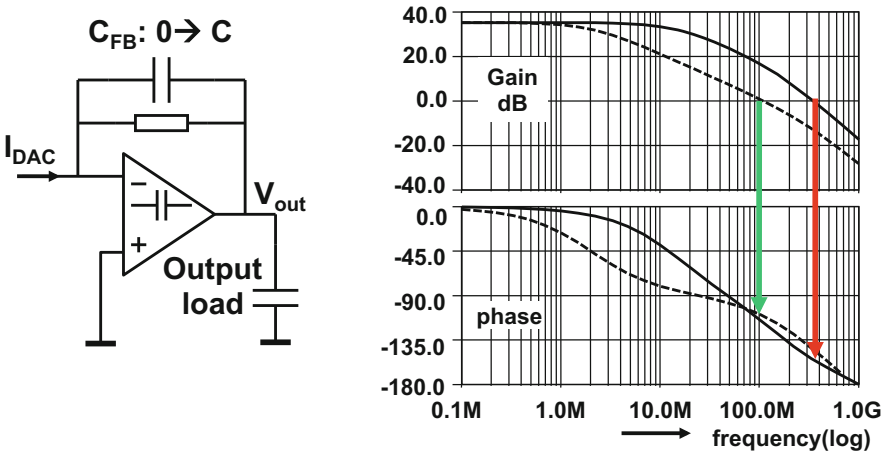


Fig. 7.28 The Miller capacitor for the current buffer must be increased to avoid instability. The *bold lines* are for zero-feedback capacitance. The *dotted lines* indicate the gain and phase for a significant feedback capacitor

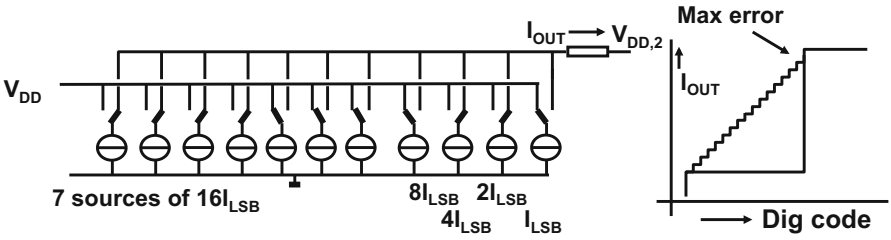


Fig. 7.29 A 3-bit unary array of current sources is complemented with a 4-bit binary array

7.3.2 Current Steering Digital-to-Analog Converter

The block diagram in Fig. 7.29 shows a digital-to-analog converter based on current sources without a bandwidth-limiting buffer. The current sources are directly feeding the load impedance. The bandwidth of this arrangement is now only limited by the pole of the output impedance. This topology is suited for delivering a high-performance time-continuous signals into 50–75 Ω loads. This architecture is called "current-steering digital-to-analog converter."

For a desired output swing and fixed load, the resulting load current must be supplied by the current sources. This implies that a current of, e.g., 15–20 mA must be distributed over the array of current sources with a sufficiently low-ohmic connection pattern.

Current-steering digital-to-analog converters use a segmented architecture: a unary section supplemented with a binary part. These converters use unary current sources for the MSB values, thereby avoiding large DNL errors for the major transitions as in a (full) binary architecture. The area penalty of unary architectures is minimized with binary-coded current sources for the lower bits. The total area is thereby considerably smaller than for a completely unary implementation. A 10-bit current steering digital-to-analog converter can, e.g., be built from 64 unary current sources and 4 binary current sources requiring considerably less area than 1024 sources for a full unary implementation. The area reduction is of the order of $2^{N_{binary}}$ where N_{binary} represents the resolution of the binary section. The circuit in Fig. 7.29 shows a 3-bit unary array supplemented by a four-bit binary array. For optimum matching the entire structure is here based on one current-source design that is combined in groups. This would not lead to the above-mentioned area reduction, so many designs sacrifice some matching by using $I_{LSB}/2, I_{LSB}/4, ..$ etc.

The choice between unary and binary block sizes depends on technology and required DNL performance [143–145]. The dominant DNL problem occurs when the binary section is switched off and replaced by the next unary source, see Fig. 7.29 (upper, right).

The INL error (for the best fitting straight line algorithm as in Fig. 4.9) at any position will be defined by all the current sources: the ones that set the signal, but also the other current sources that ultimately determine the full range. These will

affect the slope and offset of the best fitting straight line. If the accuracy of each current source is limited by random effects that are inversely related to the area, the total amount of gate area for a certain INL resolution for both unary and binary architectures is comparable. The DNL specification and secondary arguments such as wiring overhead are more relevant.

The design of the unary and binary current sources requires detailed understanding of the process, see Chap. 5. The global variations in the currents as caused by process, voltage, and temperature apply to all current sources and with careful design (see, e.g., Table 5.5) deviations can be minimized. Strong power grids are used to minimize voltage drops and guard rings and substrate rings are used to keep the local substrate clean and equipotential. However, uncorrelated effects and gradients cause errors at all transitions.

The major DNL error will occur at the transition of the binary section to the next unary current source as shown in Fig. 7.29 (upper,right) and its variance can be estimated as [145]:

$$\sigma^2_{\text{one DNL}} = \frac{\sigma^2_{I_{unary}} + \sigma^2_{I_{all\ binary}}}{I^2_{LSB}} = 2^{2N_{binary}} \frac{\sigma^2_{I_{unary}}}{I^2_{unary}} + (2^{N_{binary}} - 1)^2 \frac{\sigma^2_{I_{all\ binary}}}{I^2_{all\ binary}}$$

for a single binary to unary transition. $\sigma_{I_{unary}}$ and $\sigma_{I_{all\ binary}}$ represent the standard deviation of a single unary current source and the standard deviation of the sum of the binary current sources. With the help of Eqs. 5.16 and 5.25 the standard deviation associated with this step is

$$\sigma^2_{\text{one DNL}} = \frac{2A^2_{VT} 2^{2N_{binary}}}{(WL)_{unary}(V_{GS} - V_T)^2} + \frac{2A^2_{VT}(2^{N_{binary}} - 1)^2}{(WL)_{all\ binary}(V_{GS} - V_T)^2}$$

This equation uses the current mismatch relation 5.25 for the current sources and expresses the result in fractions of an I_{LSB}. The same is done for the total amount of binary current sources. In many designs the gate area for the binary section is comparable to the gate area for one unary source and both terms will have similar magnitudes. In that case the DNL caused by one binary to unary transition is approximately:

$$\sigma_{\text{one DNL}} \approx 2^{N_{binary}+0.5} \frac{\sigma_{I_{unary}}}{I_{unary}} \approx \frac{2A_{VT} 2^{N_{binary}}}{\sqrt{(WL)_{unary}}(V_{GS} - V_T)}$$

This equation shows that large area, a high drive voltage, and a low matching factor are the main ingredients for high performance. Based on this formula a first estimate is possible for the seize of the binary section N_{binary}.

Example 7.6. Estimate N_{binary} for a current source area of $25\,\mu\text{m}^2$, with $(V_{GS} - V_T) = 300\,\text{mV}$ and $A_{VT} = 3\,\text{mV}\mu\text{m}$.

Solution. Using the previous equation the standard deviation for the DNL is: $\sigma_{\text{one DNL}} = 0.004 \times 2^{N_{binary}}$.

For $N_{binary} = 8$ bit, the standard deviation is $\sigma_{\text{one DNL}} = 1.024\,\text{LSB}$, which means that each unary to binary transition has a chance of 34 % to exceed $|DNL| > 1$, which is by all standards too much. For 7 bit, $\sigma_{\text{one DNL}} = 0.5\,\text{LSB}$, for 6 bit $\sigma_{\text{one DNL}} = 0.25\,\text{LSB}$. So most likely not more than $N_{binary} = 5$ bit is possible.

Example 7.7. In a segmented current-steering digital-to-analog converter N_{unary}=6 bit and $N_{binary} = 5$ bit. An overall DNL $< 0.5\,\text{LSB}$ is desired for 99.73 % of the devices.

Solution. The unary matrix is composed of $2^{N_{unary}} - 1 = 63$ current sources. If at all transitions the DNL must remain below a target value $\text{DNL}_{max} < 0.5\,\text{LSB}$, with a certainty of 99.73 %, the probability for each transition to be correct must be better than the 63-rd root of 0.9973, which requires a probability of 0.999957 per transition. Table 2 shows that this probability corresponds to 4.0σ. For $N_{unary} = 6$ and $\text{DNL}_{max} < 0.5\,\text{LSB}$, $\sigma_{\text{one DNL}} < 0.125\,\text{LSB}$. Now the relative current mismatch for the unary current source (and also the sum of the binary current sources) is $2^{-N_{binary}}$ lower. In this example with $N_{binary} = 5$,

$$\frac{\sigma_{I_{unary}}}{I_{unary}} < 0.125/32 = 0.0039$$

Even with a rather high drive voltage, the area per current source transistor will be tenths of μm^2.

Example 7.8. What are the requirements on the random mismatch for a binary current-source architecture using standard current sources, and a binary current digital-to-analog converter based on an R-2R network? In both cases a 3σ yield on monotonicity is required.

Solution. For the current-source architecture:

Every current branch is composed of a parallel connection of 2^k basic current sources as in Fig. 7.26 (lower). The current branch of weight k in the array is described by $I_k = 2^k I_{LSB}$. Let each basic current source I_{LSB} suffer from a variation mechanism characterized by a normal distribution with mean value I_{LSBm} and a variance σ_I^2. Then each branch will be characterized by a mean $2^k I_{LSBm}$ and a variance $\sigma_{Ik}^2 = 2^k \sigma_I^2$, see Eq. 5.12. With the same equation the variance for the current step on the MSB transition is

$$\sigma_{\Delta I}^2 = (2^{N-1})\sigma_I^2 + (2^{N-2} + \ldots .2^1 + 2^0)\sigma_I^2 = (2^N - 1)\sigma_I^2 \tag{7.29}$$

Monotonicity requires that the value of ΔI remains positive. If a 3σ probability (99.7 %) is considered an acceptable yield loss, then $3 \times \sqrt{(2^N - 1)}\sigma_I < I_{LSB}$. For an 8-bit converter this requirement results in $\sigma_I < 0.02 I_{LSB}$.

In an R-2R ladder as in Fig. 7.18, the current splits at the first node in an MSB current I_{MSB} and a similar current for the remaining network. The impedances R_1

and R_2 are nominally equal to $2R$, but the individual resistors suffer from random variation characterized by a normal distribution with a variance $\sigma^2_{R1,2}$. With the help of Eq. 5.11

$$\begin{aligned} I_{MSB} &= \frac{R_2}{R_1 + R_2} 2^N I_{LSB} \\ \frac{\sigma^2_{IMSB}}{(2^N I_{LSB})^2} &= \left(\frac{\partial I_{MSB}}{\partial R_1}\right)^2 \sigma^2_{R1} + \left(\frac{\partial I_{MSB}}{\partial R_2}\right)^2 \sigma^2_{R2} = \frac{R_1^2 \sigma^2_{R2} + R_2^2 \sigma^2_{R1}}{(R_1 + R_2)^4} \\ &\approx \frac{2\sigma^2_{R1}}{4(R_1)^2} \end{aligned} \tag{7.30}$$

Requiring for monotonicity that $3 \times \sigma_{IMSB} < I_{LSB}$ results for 8-bit resolution in $\sigma_{R1}/R_1 = 0.002$. The ten-times higher precision that is needed for the resistor split is due to the fact that the error is determined by the combination of just two resistors, whereas the MSB in the current architecture was built from 2^{N-1} current sources.

The INL specification depends mainly on the mismatch and gradient of the unary current sources. Gradients[10] can be suppressed by means of common-centroid layout configuration, see Fig. 7.32. In rather noisy environments, also substrate currents may affect the equality. A heavy substrate-connected guard-ring enclosing the complete current matrix can equalize the substrate potential for all current sources.

The INL will be due to the unary current sources, if the errors from the binary section are kept at a low level. The derivation of the INL follows the same mathematical path as for the resistor string. When the INL is measured with respect to the best fitting straight line, the maximum value of the variance occurs in the middle, where $m = 2^{N_{unary}-1}$. Here the relative current variance is found as

$$\frac{\sigma_{I(m)}}{I_{unary}} = \frac{\sqrt{2^{N_{unary}}}}{2} \frac{\sigma_{Iunary}}{I_{unary}} \tag{7.31}$$

where it is assumed that the DNL contribution is sufficiently low. In order to find the INL of a segmented unary–binary structure, this result for the unary section must be referred to the LSB-size: $I_{unary} = 2^{N_{binary}} I_{LSB}$. As a result the standard deviation of the INL in the middle of the range is

$$\sigma_{\mathrm{INL}} = \frac{\sigma_{I(m)}}{I_{LSB}} = \left(2^{N_{unary}/2} 2^{N_{binary}} 2^{-1}\right) \frac{\sigma_{Iunary}}{I_{unary}} \tag{7.32}$$

Increasing the unary resolution with one bit means doubling the number of current sources with a $\sqrt{2}$ increase in the standard deviation describing the INL. Increasing

[10]The main remaining gradient in processes with an epitaxial layer is caused by the temperature differences in the die and voltage drops over the wiring.

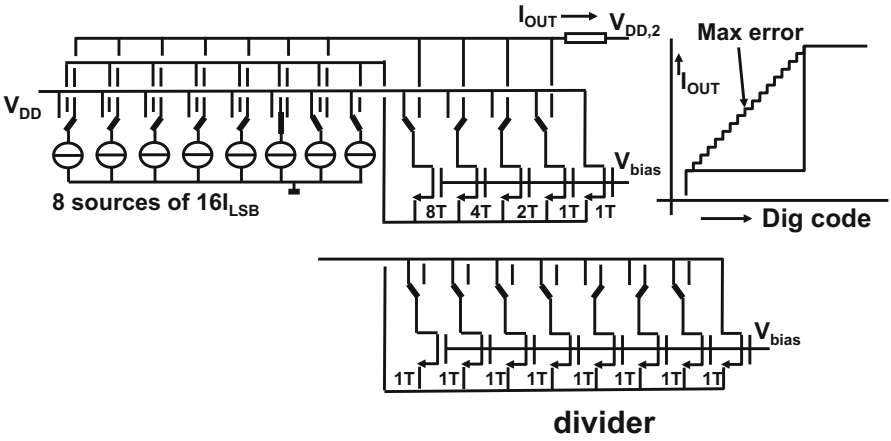


Fig. 7.30 The additional binary array is replaced by a binary [135] or unary current divider

the binary resolution with one bit results in measuring the same unary error with an I_{LSB} that is half size. So here again the difference between unary and binary is in the trade-off between area and performance. For digital-to-analog converter designs optimized on the above analysis, see [145].

An alternative implementation of a segmented unary–binary architecture is shown in Fig. 7.30 (upper) [135]. This circuit allows the unary current sources to switch to three rails. The required unary currents are switched to the output node as in the upper plot. The next unselected current source is not connected to the power supply but feeds a binary divider array. Although there are still DNL errors caused inside the binary divider, there is only a minor transition error at the transitions between the binary and unary arrays. This problem can be solved by using a full unary divider, see Fig. 7.30 (lower). In the architectures of Fig. 7.30 there is no need to match the implementation of the LSB divider section to the unary current sources. The disadvantage of this architecture is in the synchronization between the switching of the binary and unary current sources causing timing-related errors at high frequencies. Moreover there is a headroom penalty.

Example 7.9. Discuss the options to increase the resolution of a 10-bit current-steering digital-to-analog converter with $N_{unary} = 6$ bit and $N_{binary} = 4$ bit by adding 1 bit either to the unary or to the binary section.

Solution. When unary current sources with a relative random mismatch of 1 % are available (see Eq. 5.25), a structure with $N_{unary} = 6$ bit and $N_{binary} = 4$ bit will show a DNL $= 0.64$ LSB. Increasing the resolution to 11 bit by adding a bit to the unary section gives twice the area and an INL $= 0.90$ and an almost constant DNL.

If instead the binary section is expanded by one bit without much area penalty, the INL and DNL double as the same absolute error is now measured with an I_{LSB} that is half the size. Obviously the current source mismatch is a dominant design parameter.

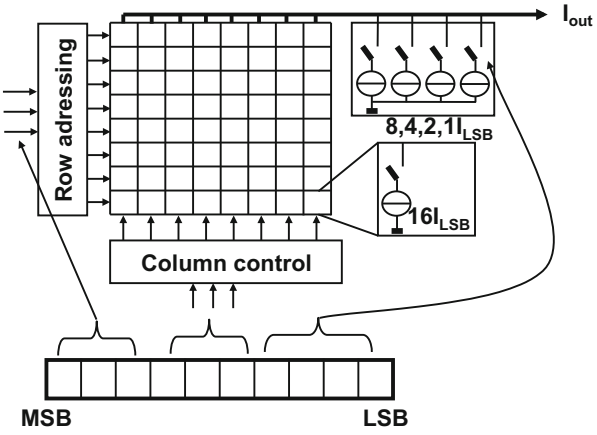


Fig. 7.31 10-bit digital–analog converter: the 6 MSBs are implemented as 64 unary current sources in a matrix configuration, while the four LSBs are designed in a binary series

7.3.3 Matrix Decoding

A popular arrangement for the digital decoding of the input word into current source control signals is a column-row addressing scheme resembling a Random-Access Memory, as indicated in Fig. 7.31. A straight-forward selection from left-to right and top-to-bottom, e.g. like someone reading this text, will emphasize gradient effects. The current sources that form the unary part of the digital-to-analog converter are preferably selected in a way that cancels gradients [133, 146, 147]. These gradients can occur due to technological deviations such as doping or oxide thickness variations, power supply drop due to the resistivity of current carrying lines, clock timing gradients, and temperature gradients. In modern epitaxial processes the technological gradients are rather limited in magnitude. Especially the voltage, time [148], and temperature gradients can become rather severe.

Figure 7.32 shows a simple solution to the first order gradient problem for an array where the current sources are connected in a binary way: the common centroid topology. Each group of current sources, forming the LSB up to the MSB, is arranged symmetrically around the center in both lateral dimensions. This will cancel any linear gradient.

For unary arrays a similar strategy can be followed. For example, start with the LSB in Fig. 7.32, then add cells from the LSB+1 group, etc.

More advanced schemes are known as Q^2-walk schemes [149]. These schemes use further subdivision of the current source array into subarrays. Groups of sub-current sources are arranged to compensate for second order components. Another approach is to randomize the current sources from one sample to the next [150].

Figure 7.33 shows a die photograph of a current-steering digital-to-analog converter based on current source sorting. An algorithm, see Fig. 7.34, first

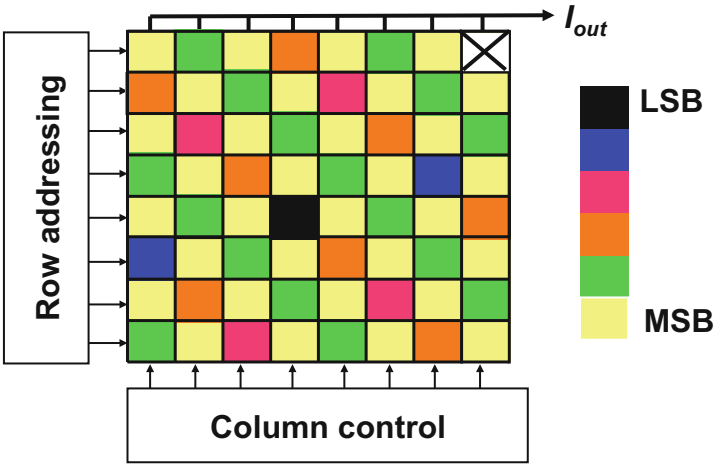


Fig. 7.32 Placement of unary weighted current sources in a 64-element array. Similar *colored squares* represent current sources that form together one bit level [147]

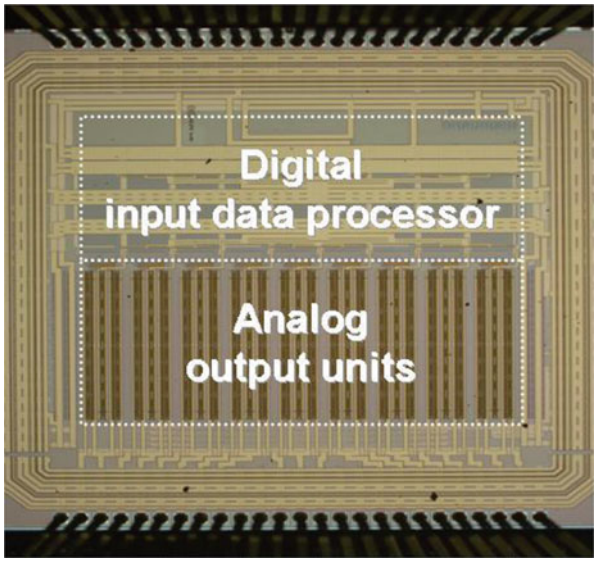


Fig. 7.33 Die photograph of a 16-bit current-steering digital-to-analog converter in 180-nm CMOS, courtesy: J. Briaire

measures the values of the current sources and then arranges them in groups that optimize the overall performance. For example, the N_{unary} unary currents are created by combining $2N_{unary}$ half-current sources in a way that one high half-current source is matched by one low half-current source. A considerable reduction of inequality can be achieved at the cost of some pre-processing.

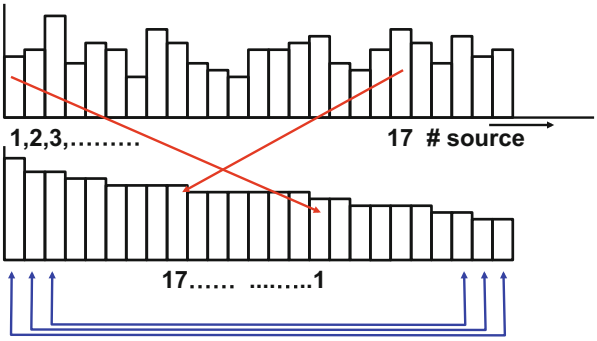


Fig. 7.34 In this sorting algorithm the current sources are measured and then pairwise matched [151]

More points to consider with matrices of current sources are

- The lines that are routed over the current-source cells can interfere with these cells due to capacitive coupling, but also in a technological way. The metal covers the current source transistors and affects the annealing of the underlying gate structure causing mismatch, Sect. 5.2.
- The decoding network consumes power and increases cross-talk.
- Close to the current switches, the selection signals must be re-timed by the main clock signal. Delays between the switching-on and off of various current sources create unequal $\int idt$ contributions to the output signal. Proper distribution of the clock signal with equal traveling distances for each cell (10 µm wire causes 0.1 ps time difference) is needed.

7.3.4 Current Cell

Figure 7.35 shows the basic current source schematic. The current source transistor is DC-biased and its size is chosen in a way to reduce mismatch effects and noise contributions (e.g., long transistor length). The current sources are switched between the output rail and the power supply. A current that is not contributing to the output current cannot be switched off. This would inevitably lead to a discharge of the inversion layers in the (rather large) current source transistors and the parasitic capacitors. Building up this charge after reconnection takes a lot of time and will lead to linear and non-linear distortion, therefore unused currents must be routed to a power rail. Consequently these converters always consume the maximum current.

One of the main issues in this architecture is the modulation of the current sources by the output voltage. This modulation is not present in the buffered current-source converters. Depending on the effective internal impedance, this voltage variation will result in some current modulation. When the digital signal or the fraction of

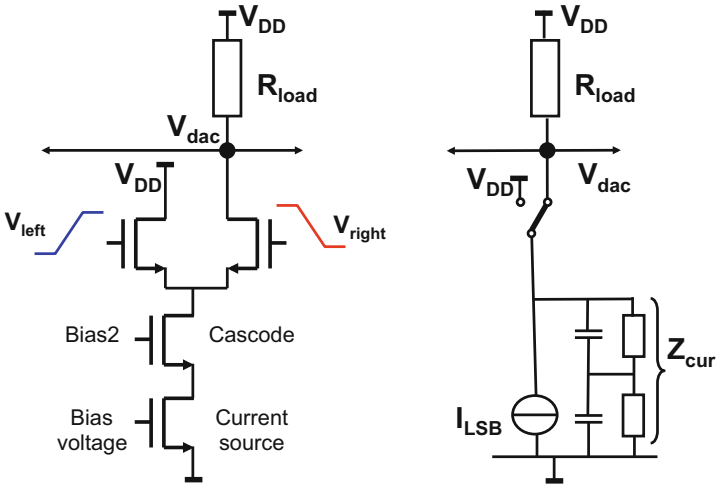


Fig. 7.35 The basic digital–analog converter current cell and its small-signal equivalent circuit

current sources that is switched on at a sample moment is modeled as $0 \leq \alpha \leq 1$, the ideal output voltage becomes

$$V_{DD} - V_{dac} = \alpha 2^N I_{LSB} R_{load} \tag{7.33}$$

where the assumption is that there are 2^N unary current sources in the unary part of the digital-to-analog converter. In fact α is treated here as an amplitude continuous signal. Its amplitude discrete nature (after all it is a digital word) is ignored, and the additional errors due to the amplitude quantization must be added later.

A finite output impedance of each active current source Z_{cur}, as shown in the right-hand side of Fig. 7.35, causes an additional current because it loads the digital-to-analog converter's output voltage. The total error current flowing into the finite output impedance of the current sources is

$$I_{err} = \alpha 2^N V_{dac} / |Z_{cur}| \tag{7.34}$$

Both the fraction of active current sources α and the output voltage V_{dac} are proportional to the signal: αV_{dac} generates non-linearity. Circuit analysis now shows

$$V_{DD} - V_{dac} = R_{load}(\alpha 2^N I_{LSB} + \frac{\alpha 2^N V_{dac}}{|Z_{cur}|}) \tag{7.35}$$

where the phase shift due to the complex impedance is ignored. Rearranging terms gives

$$V_{dac} = \frac{V_{DD} - R_{load} \alpha 2^N I_{LSB}}{1 + \alpha 2^N R_{load} / |Z_{cur}|} \tag{7.36}$$

The second term in the denominator $\alpha 2^N R_{load}/|Z_{cur}| \ll 1$ reflects the small error term, so the denominator term can be evaluated using the Taylor series expansion:

$$\frac{1}{1+\Delta} = 1 - \Delta + \Delta^2 -, \ldots.$$

For evaluation of the signal with its second and third-order non-linearity, only three terms are used and the result is

$$V_{dac} = \approx (V_{DD} - R_{load}\alpha 2^N I_{LSB}) \left[1 - \frac{\alpha 2^N R_{load}}{|Z_{cur}|} + \left(\frac{\alpha 2^N R_{load}}{|Z_{cur}|} \right)^2 \right] \tag{7.37}$$

Inspection of the main contributions in this formula shows that this is an equation of the form $V_{dac} \propto \alpha - C\alpha^2 + C^2\alpha^3$. The last two terms result in a second and third-order distortion term.

Equation 7.37 describes the transfer curve from digital signal to an analog output. Substitution of a full-swing signal $\alpha = 0.5 + 0.5\sin(\omega t)$ and applying some trigonometry: $\sin^2(\omega t) = (1 - \cos(2\omega t))/2$ results after some manipulation in:

$$\text{HD2} = \frac{\text{second order component}}{\text{first order component}} = \frac{2^N R_{load}}{4|Z_{cur}|} \tag{7.38}$$

This signal ratio can be expressed in dBs via $20^{10}\log(\text{HD2})$. The second order distortion is directly related to the ratio between the output impedance of the current sources and the load impedance.

In a differential output scheme the complementary current is used to drive a second identical load. In the differential output signal the second order distortion is largely eliminated, and only a third order component remains [152]:

$$\text{HD3} = \frac{\text{third order component}}{\text{first order component}} = \left(\frac{2^N R_{load}}{4|Z_{cur}|} \right)^2 \tag{7.39}$$

Next to the resistive modulation, Z_{cur} also contains capacitors, such as the parasitic capacitor on the drain of the current source transistor. The capacitive current will increase with higher frequencies and so will the distortion components. This effect and the time deviations of the switching pulses are the root cause for performance loss at higher signal frequencies.

Dedicated cascode stages are used for reducing the effect of output modulation. In some designs the switches are used in their conductive mode as cascode stages, however, the demands for fast switching and an output impedance increase normally conflict.

The switch transistors must be optimized for fast switching and are controlled by low-swing differential pulses [153]. This will reduce the cross-talk and avoid the creation of unnecessary inversion charge. The timing of the switches needs to be

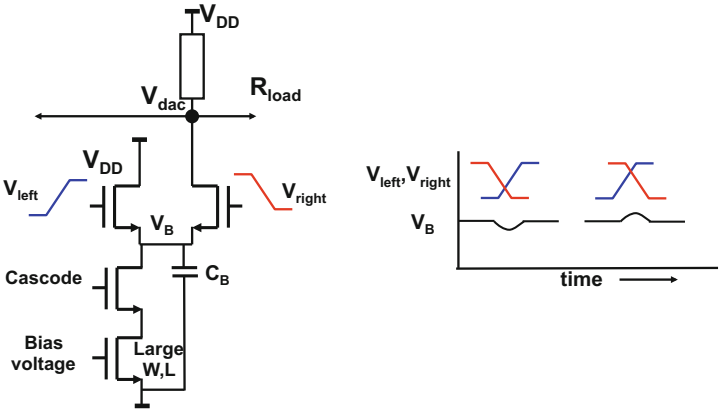


Fig. 7.36 The timing of the switch transistors is critical. Inaccurate pulses will allow charge exchange via parasitic capacitor C_B

accurate, otherwise charge will be exchanged between the branches via voltage dips or bumps on the parasitic capacitors, see Fig. 7.36. Re-timing just before the current source is often applied. Although the switches are in series with a current source, the voltage drop over the switch can affect the performance, so their on-resistance is made low. As all resistances of the current switches are in parallel, the value of $R_{switch}/2^N$ must be kept low compared to R_{load}.

Yet, the switches in current-source digital-to-analog converters complicate the design in more aspects. In contrast to the resistor ladder, where only one switch is active, in a current matrix a number of switches must be toggled if the signal experiences a larger voltage step. Each switch (MOS or bipolar) needs some charge to create a conducting path. This charge is taken from the current source and the signal during switching on, and is released into the output signal during switching off. If a current-source digital-to-analog converter is used as an output device for driving an application, this charge disturbance distorts the output signal and must be minimized. This disturbance is called a "glitch" and is harmful as this converter is used to produce a time-continuous output signal, Fig. 7.37. The glitch produces a voltage excursion over a certain time period. The associated area under the glitch is called the "glitch energy" and often expressed in picoseconds-Volt (psV). Careful design of the switches is necessary. The voltage swing on the switches must be kept at a minimum and the clock edges must be strictly timed in order to limit the glitch energy.

Other designs [154, 155, 159] try to keep the glitches constant for all codes. Figure 7.38 shows a switching method that generates an equal number of transitions for any sequence of data bits. If the digital data remains constant, the clock will cause an alternation between the two parallel branches. At each clock cycle there is one transition which reduces the variation in glitches and bumps on the joint source connection.

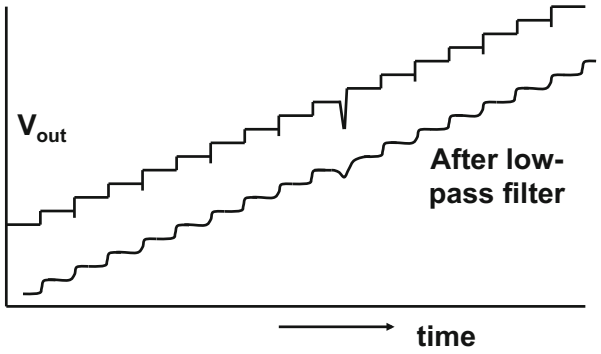


Fig. 7.37 A typical output wave form of a current-steering digital-to-analog converter. Small glitches are visible and one big glitch. After filtering it becomes clear that the glitch impacts the power contained in the signal

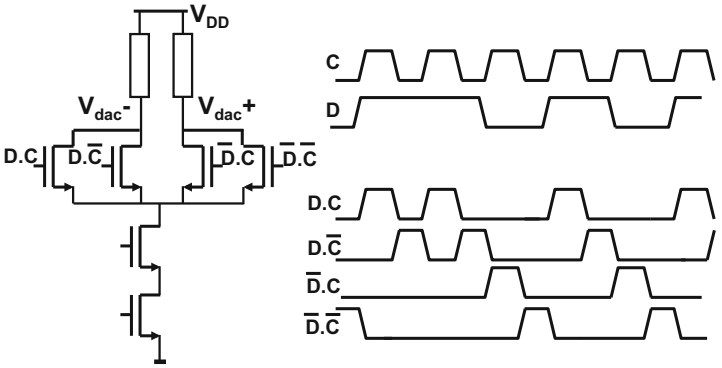


Fig. 7.38 A current cell with quadruple switches keeps the number of transitions equal for every data sequence [154, 155]

An additional problem with the switches, clock and signal lines are the timing related errors.

- Clock jitter during reconstruction of the samples in physical quantities has a similar effect as jitter in sampling. A signal frequency dependent noise contribution will appear in the output spectrum, compare Sect. 7.5.
- Wire length differences in clock lines (skew) or in current summing lines can create fixed timing errors (10 μm equals 0.1 ps). As these errors can be correlated for parts of the current sources both a noise like behavior and distortion can result.
- Mismatch in transistor parameters of clock buffers can create random variation in switching moments, see Fig. 5.46.

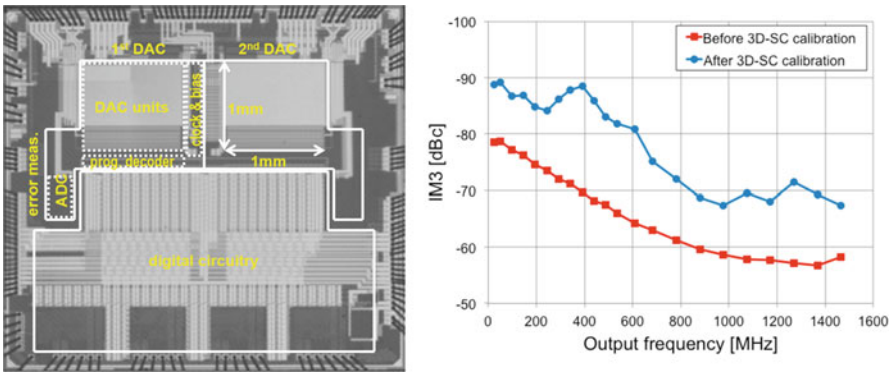


Fig. 7.39 Chip photo and intermodulation IM3 of a 16-bit current steering digital-to-analog converter [156]. The timing and mismatch errors are calibrated leading to a significant suppression of distortion. Courtesy photo and graph: Hans van de Vel, NXP

These effects create variations in the moment a current source is switched on and in the duration of the current pulse. While current source mismatch dominates the performance at low frequency, the timing effects limit the performance at high signal frequencies. Next to careful design various calibration techniques are applied to minimize the quality loss. A 16-bit converter, Fig. 7.39 [156], addresses these problems along the same line as explained in Fig. 7.34.

Example 7.10. A 10-bit unary current steering digital-to-analog converter drives a 50-Ω load. Calculate the allowable parallel rersistance of each current source for a HD2 $= -60$ dB performance. Part of the impedance is a 0.1 pF parallel capacitor. At what signal frequency will the HD2 cross the -57 dB level?

Solution. A HD2 of -60 dB means that

$$\text{HD2} = \frac{2^N R_{load}}{4|R_{cur}|} = \frac{2^{10} 50\Omega}{4|R_{cur}|} = \frac{1}{1000}$$

so $|R_{cur}| = 12.5\,\text{M}\Omega$.

At the frequency where HD2 $= -57$ dB, the RC network loses another 3 dB. This corresponds to a signal frequency where $\omega R_{cur} C_{cur} = 1$. With $R_{cur} = 12.5 \times 10^6 \Omega$, $C_{cur} = 0.1$ pF, the cross-over frequency is 130 kHz.

Note that this is not a favorable value. A reduction of the number of unary current sources and a reduction of the capacitance are needed. Going to a differential design for adequate suppression of the second-order term is sometimes unavoidable.

Example 7.11. Show that a differential configuration of a current-steering digital-to-analog converter suppresses the second order distortion.

Solution. If the constant V_{DD} term in Eq. 7.35 is ignored, the signal terms can be evaluated as

$$V_{dac} = -R_{load}\alpha 2^N I_{LSB}\frac{1}{1-\frac{\alpha 2^N R_{load}}{Z_{cur}}} \approx -R_{load}\alpha 2^N I_{LSB}(1+\frac{\alpha 2^N R_{load}}{Z_{cur}}) = C_1\alpha + C_2\alpha^2 \tag{7.40}$$

where the approximation $1/(1-a) \approx (1+a) \quad |a| \ll 1$ is used. If the complementary current $(1-\alpha)2^N I_{LSB}$ is fed in a second resistor of value R_{load}, the resulting voltage $V_{dac,inv}$ is easily found by replacing α in the equations by $(1-\alpha)$. The differential voltage is then

$$V_{dac} - V_{dac,inv} = C_1\alpha + C_2\alpha^2 - (C_1(1-\alpha) + C_2(1-\alpha)^2) = (C_1+C_2)(2\alpha - 1) \tag{7.41}$$

which is linear with the signal term α.

Example 7.12. Groups of 1, 2, 4, ... unit current sources are combined to form a binary architecture digital-to-analog converter. Each unit current source shows a random mismatch of $1\sigma = 1\,\%$ of the LSB current. How many bits can be designed when a DNL of 0.5 LSB must be reached for 99.7 % (or -3 to $+3\sigma$) of the produced devices.

Solution. The absolute mismatch of a single LSB current source is $\sigma_I = 0.01 I_{LSB}$. Putting n current sources parallel increases the average current to nI_{LSB} and the mismatch to $\sigma_{nI} = \sqrt{n}\sigma_I = \sqrt{n}0.01 I_{LSB}$. The worst-case transition in an N-bit binary digital-to-analog converter is at the switching point between the MSB current $2^{N-1}I_{LSB}$ with $\sigma_{IMSB} = \sqrt{2^{N-1}}\sigma_I$ and the LSB to MSB-1 currents: $(2^{N-1}-1)I_{LSB}$ with $\sigma_{Ilower-bits} = \sqrt{2^{N-1}-1}\sigma_I$. The nominal difference between the groups of current sources is exactly $1I_{LSB}$. The variance of the difference between these groups is

$$\sigma^2_{diff} = (2^{N-1})\sigma_I^2 + (2^{N-1}-1)\sigma_I^2 = (2^N - 1)\sigma_I^2$$

when $3\sigma_{diff} < 0.5 I_{LSB}$ the result is $N = 8$.

Example 7.13. A binary current digital-to-analog converter is built with weighted current sources, each with a random mismatch of 1 % of the current, independent of the current value. How many bits can be designed as a binary section when a DNL of 0.5 LSB must be reached for 99.7 % (or -3 to $+3\sigma$) of the produced devices.

Solution. The absolute mismatch of the i-th current source ($i = 0, .., N-1$) is now $0.01 I_{LSB} \times 2^i$. The worst-case transition in an N-bit binary digital-to-analog converter is at the switching point between the MSB current $2^{N-1}I_{LSB}$ with $\sigma_{IMSB} = 2^{N-1}0.01 I_{LSB}$ and the LSB to MSB-1 currents: $\sigma_{Ilower-bits} = 0.01 I_{LSB}\sqrt{2^{2(N-2)} + \ldots + 2^8 + 2^4 + 2^2 + 2^0}$. The nominal difference between the groups of current sources is $1I_{LSB}$. The variance of the difference between these groups is

$$\sigma^2_{diff} = 2^{2(N-1)}(0.01 I_{LSB})^2 + (2^{2(N-2)} + \ldots + 2^8 + 2^4 + 2^2 + 2^0)(0.01 I_{LSB})^2$$

when $3\sigma_{diff} < 0.5I_{LSB}$ the result is $N = 3$. The 1 % relative spread in the MSB current clearly dominates. If a process shows large sensitivity to relative spreads (e.g. random mask variations in the gate length), it is better to design the converter with small current sources.

7.3.5 *Performance Limits*

Current-steering converters without speed-limiting buffer reach high operating frequencies. The bandwidth constraint at the output of these converters is determined by the time constant of the output load, which also serves as a first alias filter. The penalty for these advantages is power. A single terminated 50 Ω cable requires to supply 20 mA per converter. An advantage of the constant current consumption is that the impact of power wiring impedances and bond-wire inductance is low as (in first order) there is no signal dependent current flowing through these wires. Next to the current sources also the high-frequency clock and decoding schemes require a significant amount of power. Moreover the speed of the digital processing and the timing inaccuracy limit the performance.

In differential systems the drained current is used to implement a differential output that will cancel the second order distortion component.

Figure 7.40 compares data from various publications [144, 148, 150, 155–159] and data sheets on current-steering digital-to-analog converters in the time frame 2000–2015. The plot suggests a first-order relation between the bandwidth and the level of spurious and harmonics. Another metric of comparison is a Figure of Merit, see Eq. 4.27.

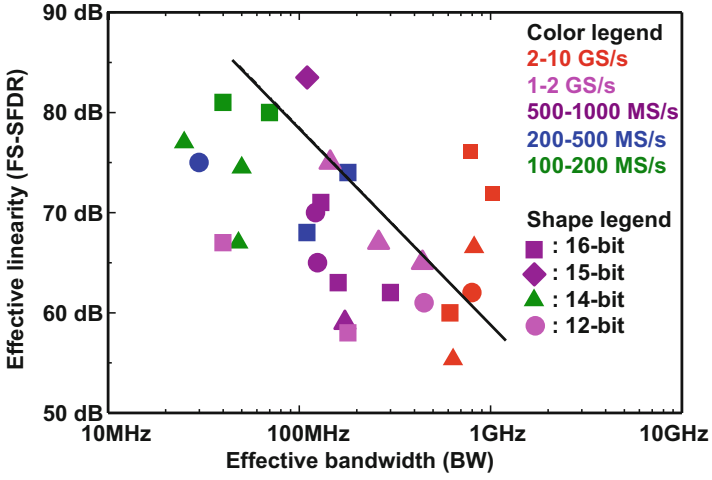


Fig. 7.40 Spurious-free dynamic range versus bandwidth for current-steering digital-to-analog converters published from 2000–2015. Based on a plot by: J. Briaire

Current-steering digital-to-analog converters are the industry's primary choice for output devices in demanding applications. Their dynamic performance is unmatched and the power penalty is accepted.

Example 7.14. A unary current matrix with 1024 current sources produces a maximum current of 20 mA in a 50 Ω load. The current sources are built in 0.18 µm CMOS with an output impedance of 100 kΩ. A single-transistor cascode stage is used. What is the total output impedance of one current source if the gate cascade transistor measures 1 µm/0.18 µm? What will be the distortion (THD) if a maximum amplitude sine wave is applied. What must be done to reduce the distortion to −60 dB.

If a parasitic capacitance of 100 fF is present parallel to each current source, what will be the frequency where the distortion is raised by 3 dB?

Solution. A minimum gate-length transistor has according to Table 3.1 a static feedback factor of $\lambda = 0.05$. So this cascode stage increases the output impedance of the current source to $Z_{cur} = r_0(1 + 1/\lambda) = 100\,\text{k}\Omega(1 + 20) = 2.1\,\text{M}\Omega$, see Table 3. The dominant distortion component is the second order term:

$$\text{HD2} = \frac{2^N R_{load}}{4Z_{cur}} = 6.1 \times 10^{-3}$$

or 44.3 dB.

In case of $N = 10$, $R_{load} = 50\,\Omega$ and a desired −60 dB level of the second order distortion HD2, the effective impedance of a single current source must be better than 12.8 MΩ which can be achieved by choosing longer transistors or adding another cascode stage.

If the distortion increases by 3 dB, the impedance $|Z_{cur}|$ must be 3 dB lower. As Z_{cur} is formed by the resistive output impedance of the cascode stage and the loading capacitor of 100 fF, $|Z_{cur}|$ will drop 3 dB at the frequency where the capacitive impedance equals the resistive impedance:$r_0(1 + 1/\lambda) = 1/(2\pi fC)$. This results in $f_{3db} = 0.76$ MHz.

7.3.6 *Comparison Between Current Source and Resistor String Digital-to-Analog Conversion*

Current-steering digital-to-analog converters designed with a current-cell matrix allow fast and accurate settling. Resistor-string converters as in Sect. 7.2.7 run at much lower power and have inherently good DNL properties. This section compares both approaches.

For this comparison a 100 Ms/s 10-bit digital-to-analog converter was designed with an array of 64 current sources for the 6 MSBs. This structure was completed with a 4-bit binary-coded LSB section. Figure 7.31 shows the block diagram of the 10-bit design. The major design issue in this circuit is the switching of the 64

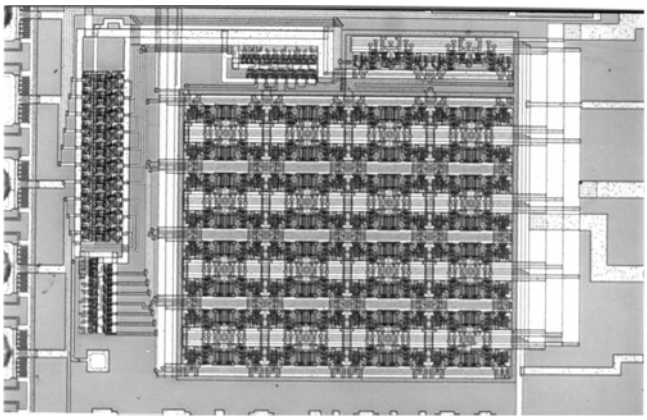


Fig. 7.41 Good old days (1990, 1 μm technology, two metal layers). Chip photograph of a 10-bit current-steering digital-to-analog converter. Digital input buffers are to the *upper left*, the horizontal decoding is below, above the unary 8 × 8matrix to the *left* is the vertical decoder and to the *right* the binary array

unary current sources. Decoding delays in the 64 current sources are reduced by an additional latching stage just before the current switch. The current switch itself is designed with a low-swing differential pair. This results in a low clock feed-through on the output line, while in this case the switch transistors also act as a cascode stage, thus reducing the modulation of the output current by the output voltage. Figure 7.41 shows the chip photograph of this design.

Table 7.4 compares the specifications of the two 10-bit video digital-to-analog converters. Remarkable differences are the differential linearity error and the distortion. The DNL error in the current-steering digital-to-analog is a direct effect of the current source mismatch: especially between the unary and binary sections. In the voltage digital-to-analog this problem is circumvented by a fully unary approach: 1024 resistors in series are used. The consequences are of course seen in a larger area.

The harmonic distortion has different origins in the two converters: in current-steering digital-to-analog converters the non-linear behavior of the large output diffusion node and the output transconductance of the current sources is important. The distortion in the voltage digital-to-analog converter is caused by the limited performance of the driver stage in the output buffer. The modulation of the switch resistance in the resistive digital-to-analog is effectively canceled by the ladder organization, while the reduced swing scheme of the current source switching limits the switch distortion in the current digital-to-analog converter.

The dynamic behavior of the digital-to-analog converter is determined by the output pole: in the current digital-to-analog converter this is the dominant pole with a 25 pF/75 Ω load. The opamp that implements the buffer of the resistive digital-to-analog converter sees this pole as its second pole because the internal

Table 7.4 Comparison of measured specifications of a ladder and a current digital-to-analog converter, both loaded with 75 Ω and 25 pF

Digital-to-analog converter type	Ladder	Current
Process	1.0 µm CMOS	
DC resolution	10 bit	
Sample frequency	>100 MHz	
Area 1 µm CMOS	1.05 mm^2	0.7 mm^2
Differential linearity error	<0.1 LSB	<0.6 LSB
Integral linearity error	<0.35 LSB	<1 LSB
Glitch energy	100 psV	100 psV
Rise/fall time (10–90 %)	4 ns	1 ns
Settling time (1 LSB)	20 ns	5 ns
Signal bandwidth (−1 dB)	20 MHz	>20 MHz
Minimum power supply (THD< − 40 dB)	3 V	3 V
Output in 75 Ω	1 V	1 V
Output at min. code	100 mV	<1 mV
Av. Current (50 MHz, 75 Ω)	10 mA	15 mA
Av. Current (50 MHz, 2 × 75Ω)	10 mA	28 mA
THD f_{signal} = 1 MHz, f_{clock} = 27 Ms/s	−58 dB	−60 dB
THD f_{signal} = 5 MHz, f_{clock} = 100 Ms/s	−50 dB	−44 dB

Miller-compensation is the dominant pole. The buffered digital-to-analog converter output is consequently slower than the current-steering output, which is seen in differences in the rise/fall time.

The minimum value for the output code is significantly larger than 0 V for a buffered output because a minimum saturation voltage is needed in the output stage.

Voltage domain digital-to-analog converters route about 75 % of their average current into the 75 Ω output load; with those (system-determined) output loads the potential for further power reduction seems to be low on an implementation level. The difference in power dissipation is less pronounced because there is more overhead for ladders and biasing in a voltage domain. Even so, a factor of 1.5 to 2.8 remains.

If this comparison between both digital-to-analog converters is repeated in modern processes, most of the above arguments will remain valid. However, the absolute speed performance will easily go up by a factor 10.

7.3.7 Semi-Digital Filter/Converters

After the conversion from the digital to the analog domain with a low-resolution digital-to-analog converter, the higher alias bands contain a lot of quantization power, despite some filtering by the zero-order hold function. If the digital signal

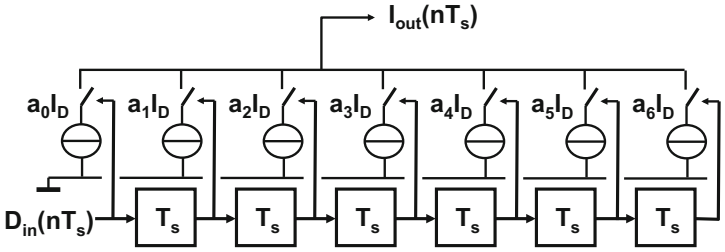


Fig. 7.42 A semi-digital filter and digital–analog converter [160]

is a high-frequency 1-bit data stream, see, e.g., Sects. 10.2 and 10.3, most of the energy in the sampled data stream is quantization error energy. In many applications the high-frequency energy must be removed before the signal can be used.[11] First converting the signal in its analog form with subsequent filtering is possible, but every non-linearity in that path will directly lead to intermodulation between the high-frequency components. With many mixing products in the low-frequency range as a result. Digital filtering is feasible, but often expensive: in area, in delay, in power.

Su and Wooley [160] proposes a digital-to-analog structure that converts the digital signal back into its analog form that at the same time filters the result from unwanted high frequency components. The semi-digital filter in Fig. 7.42 shifts the signal through a digital shift register. Every output of the shift-register sections controls a weighted current source. The relative weight of the current source is now the multiplication factor and the outputs of these current sources are simply summed by connecting them together.

The most simple implementation uses a one-bit digital signal that will switch on or off the weighted current sources. With the help of Sect. 2.7.1 the result is written as:

$$I_{out}(nT_s) = \sum_{k=0}^{k=K-1} a_k D_{in}((n-k)T_s) I_D \tag{7.42}$$

The coefficients are chosen using similar constraints as for a normal FIR filter. This structure is very useful in the reconstruction of delta-modulated signals [161].

The advantage of this structure is the suppression of the alias components in the current domain. Voltages will only appear after the current summation. So higher order signals components can be effectively cancelled in the current domain and do not generate distortion when they appear as voltages over non-linear components.

It is important to realize that errors in the coefficients in first order will affect the filter characteristics and not the overall linearity. An error in the current sources

[11] Think of all the energy your tweeter loudspeakers would have to consume.

of a unary digital-to-analog converter will cause a linearity error at that level with harmonic distortion as a result. In a semi-digital converter this error modifies the weighting coefficient by Δa_i and will result in an additional term of $1 + \Delta a_i z^{-i}$ which is a linear time-shifted signal contribution. The filtering will be less perfect but no harmonic distortion is generated.

7.4 Digital-to-Analog Conversion in the Charge Domain

7.4.1 Switched Capacitor Digital-to-Analog Conversion

Charge-domain converters differ from voltage-domain converters because the charge is now the information carrying quantity. Even in the presence of perfectly linear capacitors, this distinction is still relevant, as e.g. offset-voltages can have different effects in the charge or voltage domains.

A stray-insensitive switching topology uses a standard switched capacitor technique to move charges. The parasitic capacitances connected to the switched capacitors in Fig. 7.43 are either charged from voltage sources or do not see any charge change because they are connected to the (virtual) ground. In Fig. 7.43 the unit capacitors are grouped in a binary ascending scheme: $2^0C, 2^1C, 2^2C \ldots .2^{N-1}C$. Only one switch per group is necessary to implement a binary coded converter.

A closer look at this structure shows that the switches on the right-hand side of the capacitor can be removed or replaced by, e.g., a reset switch. In this digital-to-analog converter the digital data is converted in an analog value by bottom-plate switched capacitors, see Fig. 7.44. In [162] this structure was proposed in an analog-to-digital converter. An early example of a full-differential binary weighted implementation is found in [163]. Unary, binary, and segmented conversions can be realized, depending on the choice of capacitor values. If in Fig. 7.44 m unit

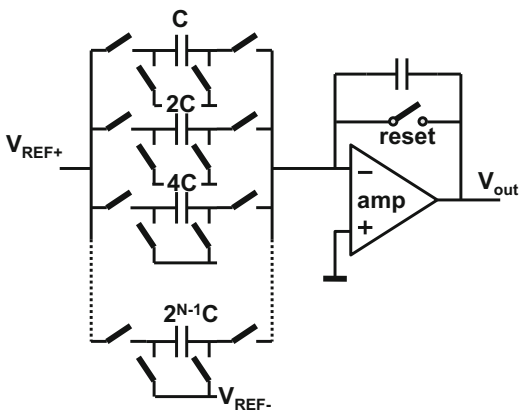


Fig. 7.43 Binary-weighted digital-to-analog converter in a stray-insensitive switched capacitor configuration

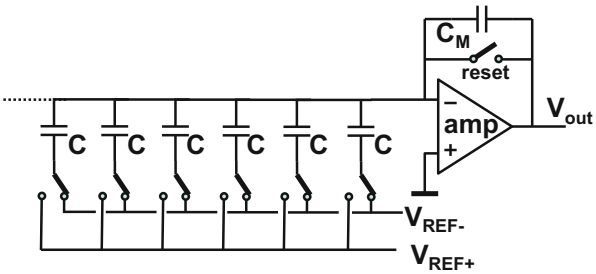


Fig. 7.44 Digital–analog converter based on unary capacitors

capacitors switch from V_{REF-} to V_{REF+}, a total charge of $mC(V_{REF+} - V_{REF-})$ is shifted in to C_M. The output voltage will now be

$$V_{out} = -\frac{mC(V_{REF-} - V_{REF+})}{C_M}$$

The configuration in Fig. 7.44 does not suffer from parasitics that are attached to the top plates of the capacitors as the virtual ground of the opamp stabilizes this node. High-speed performance requires to scale the switches to the capacitors, in order to obtain equal charging time constants. The low-frequency performance of this topology suffers most from the inequality of the capacitor banks at higher resolutions. However with capacitors that can achieve a matching performance for 12–14 bit accuracy, the more practical limitation is in the exponential growth of the capacitor bank area.

In simple capacitor arrays, the INL and DNL follow the same pattern as in current-steering arrays and resistor ladders. If the relative standard deviation of a single unit capacitor is given by σ_C/C than for a group of m parallel connected capacitors, the average value and standard deviation are

$$E(C(m)) = mC, \quad \frac{\sigma_{Cm}}{C} = \frac{\sigma_C\sqrt{m}}{C}$$

In a unary architecture with 2^N capacitors, the DNL is given by the variation of each newly added capacitor. The nominal value C corresponds to an LSB and a deviation ΔC gives therefore a DNL error of LSB$\times\Delta C/C$. In the DNL curve this error can be partitioned in a positive and negative portion, which cosmetically halves the number. The DNL curve is based on 2^N instances of σ_C/C. Even for very small capacitors $\sigma_C/C < 1\%$ and the maximum DNL remains below $\ll$0.1 LSB.

In a properly laid-out design the INL curve is also determined by the random variation in the capacitor array, and follows the same procedure as for a resistor string in Eq. 7.25:

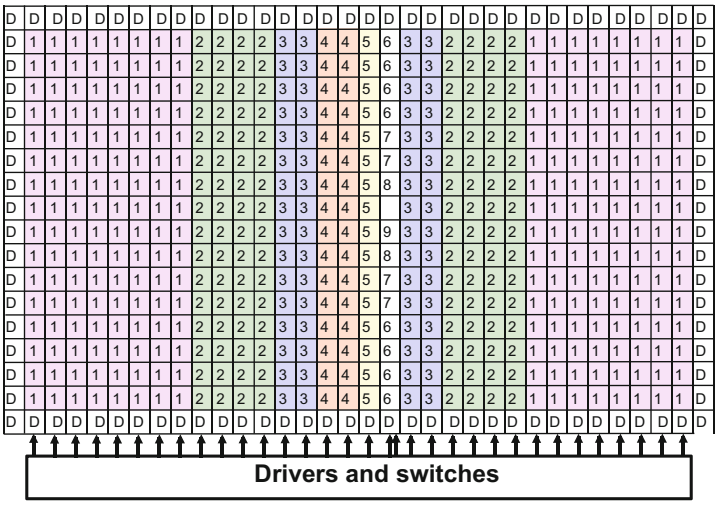


Fig. 7.45 This capacitor array with driver structure will support a 9-bit full-binary (drawn) or segmented architecture. The 31 columns labelled 1–5 can be driven successively resulting in a 5-bit unary and 4-bit binary operation. It is wise to alternate between left and right side. If the drivers of all columns "1" are connected together, and the same is done for column drivers "2",. . . ,"5" a full binary architecture arises. Many variants are possible. Some designs [164] place the 7, 8, and 9th bit at the lower end of the array, to save wiring distance

$$\sigma_{INL}(m) = \sqrt{\frac{m(2^N - m)}{2^N}} \frac{\sigma_C}{C}$$

A unary capacitor array needs a switch for every capacitor. The amount of wiring involved and the sensitivity to parasitic capacitances of the rather small (1–5 fF) unit capacitors makes that a pure unary topology is avoided.

In a binary capacitor architecture, all capacitors belonging to the same binary group are tied together, thereby minimizing the wiring overhead. A popular array topology is shown in Fig. 7.45 [164]. The potential gradients are canceled as the structure is in-line symmetrical. In [164] the capacitors for the lowest bits are placed at the driver side of the array, thereby minimizing the wiring capacitance at the cost of a minor asymmetry. The drivers are placed in a row, where every column of cells (except for the 6th to the 9th bit) has its own driver. The corresponding time-constants are comparable and the switching delay will not be code dependent. This structure can be simply modified to a segmented structure: the columns act as the unary blocks, while the columns with the 6th,..,9th bit are operated in a binary switching sequence. The basic unit cell is constructed by surrounding the top plate completely by the bottom plate, Fig. 7.46 [164]. This structure can be made more area efficient by stacking more layers on top of the shown three layers. An alternative is shown in [165].

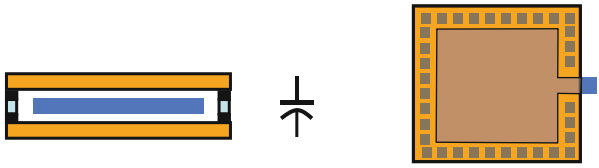


Fig. 7.46 Cross-section and top lay-out view of a unit capacitor [164]. The *top plate* of the capacitor is completely surrounded by the *bottom plate*, thereby minimizing effects of stray fields. The connection terminal should be kept at minimum size. The structure can be extended by stacking more layers on top of each other

In a binary architecture at the transition of bit i ($i = 0, 1, \ldots, N-1$) a group of $2^i - 1$ capacitors is replaced by a group of 2^i capacitors. The statistical properties of the difference are in first order:

$$E(\Delta C_i) = E(2^i C - (2^i - 1)C) = C, \tag{7.43}$$

$$\frac{\sigma_{\Delta C_i}}{C} = \frac{\sigma_C \sqrt{2^i + 2^i - 1}}{C} = \frac{\sigma_C}{C}\sqrt{2^{i+1} - 1} \tag{7.44}$$

Obviously at the major binary transition, the full variation between the capacitors forming the MSB and the capacitors forming the lower bits becomes visible. For a 10-bit binary array, 511 capacitors are replaced by 512 other capacitors and the standard deviation for the MSB is $\sqrt{1023}\sigma_C/C$, $\approx$ 32× the unary σ_C/C. With $\sigma_C/C = 1\%$ the error at the transition is $\sigma_{\Delta C_i}/C \approx 0.32$. For 3σ and spreading the error in a positive and negative part, the converter would reach DNL $= \pm 0.48$ LSB.

7.4.2 Bridge Capacitor

The span of capacitance values in digital-to-analog converters based on binary capacitor arrays requires some trade-off for the capacitor value. In some situations the choice for the unit capacitor and the lowest capacitor value is determined by the kT/C noise of the total array. However also the technological realization of small capacitors can be an issue: it is technologically difficult to fabricate accurately a capacitor far below a 1 fF value. And it is very hard to keep the parasitics under control. Figure 7.47 shows two options to extend the LSB value. A series connection of two minimum size capacitors result in half the value. As this is the LSB, an accuracy of 10 % is sufficient. Another option involves designing intermediate voltages for the LSB capacitor. This can be implemented via a resistor string, which will consume power, or via capacitive division via a bridging scheme.

The bridging capacitor, see Fig. 7.48 is an elegant solution to implement the voltage division. In this example the three MSB bits are formed in a conventional binary fashion. The three LSB bits use a bank with the same size capacitors, but are coupled via a bridging capacitor C_{bridge} to the MSB side.

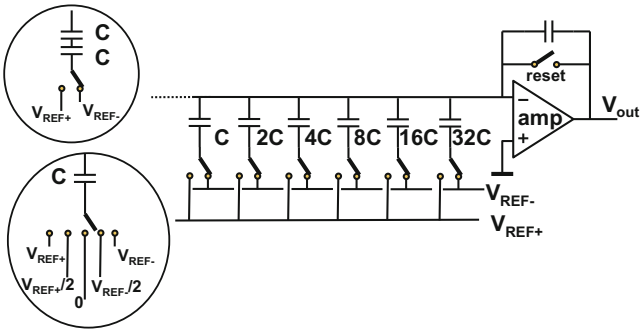


Fig. 7.47 Digital–analog converter with LSB extensions

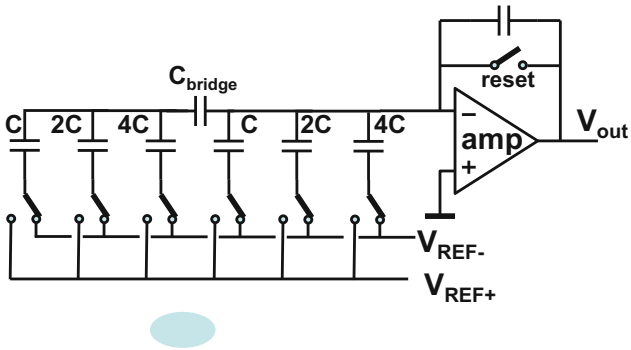


Fig. 7.48 Digital–analog converter with bridging capacitor

The capacitance of an LSB cell equals C, the LSB+1 cell equals $2C$ and the largest cell uses $2^{k-1}C$. The total capacitance on the LSB side is $(2^k - 1)C$ If an amount of iC on the LSB side is switched from negative reference to positive reference while the remaining capacitance $(2^k - 1 - i)C$ stays connected to the negative reference, the voltage change on the left-hand side of C_{bridge} is

$$\Delta V_{left} = \frac{iC}{C(2^k - 1) + C_{bridge}}(V_{ref+} - V_{ref-}) \tag{7.45}$$

Intuitively $C_{bridge} = C$ results in an exact voltage division. Formally this charge injection in the integration capacitor is compared to the desired charge injection in the summation node of the opamp:

$$\frac{C_{bridge}}{C_{bridge} + (2^k - 1)C} iC(V_{ref+} - V_{ref-}) = \frac{i}{2^k} C(V_{ref+} - V_{ref-}) \tag{7.46}$$

with the same result. The bridge capacitor can be understood as the capacitive implementation of Fig. 7.47.

Note that parasitic capacitances connected to the left side of the bridge capacitor will affect the performance as they attenuate the ΔV_{left} signal.

In first instance the capacitive schemes seem a one-to-one copy of the resistor schemes in Fig. 7.12. Yet there are some important differences. In the charge domain no constant current flow is required, except for the buffer. So a good power efficiency is possible. A second difference is in the way jitter influences the output. If time uncertainty occurs at the switching moments, the transfer of charge will be a bit early or late, however the total magnitude of the packet remains intact. In a voltage or current domain digital-to-analog converter, the overall packet consists of the time period multiplied by the current or voltage amplitude. The jitter thereby changes the signal and jitter is directly translated into noise, see also Sect. 7.5 and Fig. 10.55.

7.4.3 Algorithmic Digital-to-Analog Converter

Translation of current and voltage domain topologies to the charge domain is possible. The charm of the charge domain is, however, in the observation that storage is for free. A large number of interesting algorithms are based on this property. A simple charge redistribution digital-to-analog converter is shown in Fig. 7.49. This converter operates on a sequential binary principle: every bit is evaluated successively. After the reset switch has discharged capacitor C_2, the sequence can start. First the LSB value decides whether C_1 is charged to the positive (via S_{11}) or negative (via S_{10}) reference. Then switch S_2 connects the capacitors in parallel, redistributing the charge from the reference over both capacitors, thereby halving the value. If these capacitors are equal, half of the charge is in either. Now the LSB+1 bit is used to charge C_1, thereby destroying the remaining charge. The sequence continues with a charge redistribution, where the new charge is added to the previous charge and halved.

$$V_{out}(i) = b_i \frac{C_1}{C_1 + C_2} V_{ref} + \frac{C_2}{C_1 + C_2} V_{out}(i-1) \tag{7.47}$$

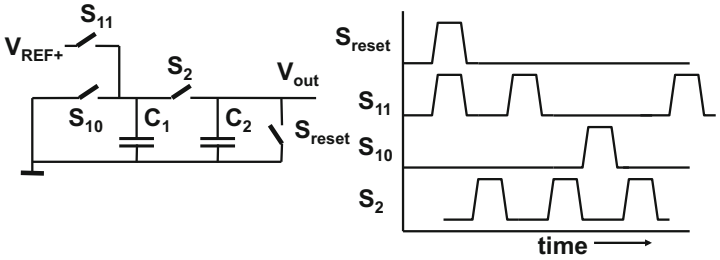


Fig. 7.49 A charge redistribution converter [162]

where $b_i \in \{0, 1\}$, with $i = 0, \ldots, (N-1)$. If both capacitors are equal, this principle is only limited by the accumulated kT/C noise and will result after N sequences in an output value of:

$$V_{out}(N) = \sum_{i=0}^{i=N-1} \frac{b_i 2^i}{2^N} V_{ref} \tag{7.48}$$

where b_0 is the LSB and b_{N-1} is the MSB. Using $\delta = (C_2 - C_1)/(C_1 + C_2)$ to describe the error between both capacitors, an estimate of the overall error in V_{out} can be made:

$$\delta V_{out}(N) = \sum_{i=0}^{i=N-1} \frac{b_i 2^i \delta^{N-1-i}}{2^N} V_{ref} < \delta V_{ref} \tag{7.49}$$

Assuming that the error at the MSB transition must remain under the value of an LSB, sets a limit for the capacitor deviation:

$$\delta = \frac{C_2 - C_1}{C_1 + C_2} < 2^{-N} \tag{7.50}$$

An error of 0.1 % in the capacitors limits the achievable resolution to $N = -^2 \log(0.001) \approx 10$ bits.

These switched-capacitor digital-to-analog converters are mostly applied in larger system chips, where they perform low-power conversions.

Example 7.15. Figure 7.50 shows a variant on the design of Fig. 7.49. The incoming bits are added to the stored result which is then amplified by 2, and stored for the next cycle in the capacitors. Now the operation starts with the MSB that is fed into the algorithmic digital-to-analog converter, processed multiplied by 2 and followed by the MSB-1. Discuss the merits of this modification.

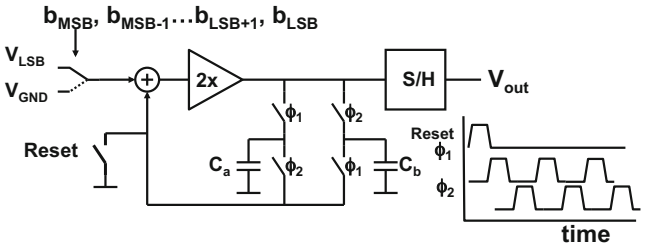


Fig. 7.50 An alternative implementation of an algorithmic digital-to-analog converter

Solution. The result of this operation bears great resemblance to Eq. 7.48: $V_{out}(N) = \sum_{i=N}^{i=1} b_{i-1} 2^i V_{ref}$, where b_0 is the LSB and b_{N-1} is the MSB. The denominator term 2^N is missing. In order to keep the overall result of the addition within operation margins, V_{ref} must now equal the small value of V_{LSB}. Any error on this quality (noise, etc.) will be amplified by the loop and result in poor signal performance.

In a preferred implementation the multiplication is with a factor 0.5, the reference value is the V_{ref} and the sequence starts with the LSB. Due to the successive divisions in that scheme, any error that occurred in the first cycles with respect to the large V_{ref} is further reduced.

Example 7.16. Charge domain digital-to-analog converters can suffer from capacitor mismatch and from KT/C noise. At what signal level is the contribution of both effects in terms of energy equal for a 1 pF capacitor?

Solution. A 1 pF capacitor will show a mismatch with respect to another capacitor. According to Eq. 5.32: $\sigma_{\Delta C}/C = A_C/\sqrt{C}$. With the value in Table 5.2 and C = 1000 fF, the relative mismatch is found as: 1.6×10^{-4} or 0.16 fF. This same signal-to-spurious ratio can be expected for a processed signal A_{rms}, resulting in an rms error of $1.6 \times 10^{-4} A_{rms}$.

In Fig. 2.21 the kT/C noise equals 65 μV_{rms}. Both contributions are equal if $1.6 \times 10^{-4} \times A_{rms} = 65 \times 10^{-6}$ V_{rms}, or $A_{rms} = 0.4$ V_{rms}. This value corresponds to, e.g., a sine wave of 0.56 V amplitude or 1.12 $V_{peak-peak}$. It is clear that both sources of unwanted energy play a different role in a design. Note that, while kT/C noise acts as a noise floor from 0 to $f_s/2$, the effects of capacitor mismatch can take many forms: from spurious single-tone component to shaped noise after data-weighted averaging.

7.4.4 Diophantic Digital-to-Analog Converter

Algorithmic principles [166, 167] allow to realize digital-to-analog converters for high-resolution application in CMOS switched-capacitor technology. The major problem which has to be solved in many high-resolution converters based on switched-capacitor techniques is performance loss due to opamp offset and mismatching in capacitor values. This digital-to-analog converter is based on an algorithm which ensures monotonicity despite offset and capacitor mismatch.

For the implementation of an N-bits converter the basic digital-to-analog conversion formula is rewritten:

$$V_{out} = V_{LSB}(a_0 2^0 + a_1 2^1 + \ldots + a_p 2^p + \ldots + a_{N-1} 2^{N-1}) = V_{LSB}(W_{LSB} + W_{MSB} 2^p) \tag{7.51}$$

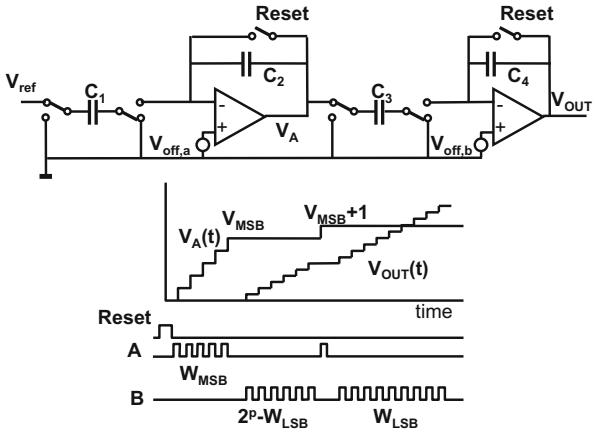


Fig. 7.51 Basic switched capacitor network with the timing diagram for converting "1011010" [166]

where W_{LSB} denotes the digital word formed by the p least significant bits: $(a_0 \ldots . a_{p-1})$ and W_{MSB} is the digital word formed by the $(N-p)$ most significant bits $(a_p \ldots . a_{N-1})$. Under the constraint that coefficients a_i, W_{LSB} and W_{MSB} are non-negative integers this equation belongs to the class of "diophantic equations." Diophantic equations are composed of polynomials with only integer coefficients and integer variables. Direct implementation in CMOS switched capacitor technique could be done in a two-stage schematic as in Fig. 7.51 by generating the analog value of an LSB at the output of section A. This V_{LSB} is W_{LSB} times transferred to section B: $V_{out} = V_{LSB} \times W_{LSB}$. Now the output of section A is raised to $2^p V_{LSB}$ and W_{MSB} transfers take place to section B. Resulting in: $V_{out} = (W_{LSB} + 2^p W_{MSB}) V_{LSB}$. This implementation leads, however, to a signal-dependent offset at the output. The offset of section B $V_{off,B}$ is transferred $(W_{MSB} + W_{LSB})$ times into the output signal:

$$V_{out} = (W_{LSB} + 2^p W_{MSB}) V_{LSB} + (W_{MSB} + W_{LSB}) V_{off,B} \tag{7.52}$$

thereby creating a non-linear dependence and signal distortion.

The contribution of the offset of section B at the output is made constant by keeping the number of charge transfers of section B constant for the conversion of any code. This can be obtained by rewriting the last part of Eq. 7.51 in:

$$V_{out} = V_{ref} \frac{C_1 C_3}{C_2 C_4} (W_{MSB}(2^p - W_{LSB}) + (W_{MSB} + 1) W_{LSB}) \tag{7.53}$$

If M_1 through M_4 are the sequential numbers of transfers, where M_1 and M_3 denote the number of transfers of section A and M_2 and M_4 are the transfers of section B, then the choice:

$$M_1 = W_{MSB},$$
$$M_2 = 2^p - W_{LSB},$$
$$M_3 = 1,$$
$$M_4 = W_{LSB}$$

results in $M_2 + M_4 = 2^p$ transfers of the offset of section B, which is no longer signal dependent. In the lower part of Fig. 7.51 the transfers for $N = 7$ and $p = 4$ have been indicated. Equation 7.53 can be interpreted as an interpolation algorithm: section A forms voltages which are proportional to W_{MSB} and $W_{MSB} + 1$, while section B interpolates in 2^p cycles between these values according to the value of W_{LSB}. The maximum number of charge transfers is $2^p + 2^{N-p}$. The fixed offset of section A is added to V_{LSB}, resulting in a gain error, and the offset of section B is multiplied by $2^p C_3/C_4$ which is signal independent and added to the output voltage. The capacitor ratios form a multiplication factor for the complete conversion, thereby influencing only the absolute gain. With a capacitor ratio of 2^{-p} a maximum swing at all opamp outputs is obtained.

The principle described by Eq. 7.53 has been the basis of the implementation of a 15-bit CMOS digital-to-analog converter with three sections of 5 bits. During the first half of the conversion period, Eq. 7.53 is used to form the value represented by the 10 most significant bits, then Eq. 7.53 is once more used to obtain the resolution for the five least significant bits.

Two important phenomena that influence the performance of the device are the limited DC gain and the settling of the charge transfer. In switched capacitor filter applications the finite gain is merely a constant factor for the transfer characteristics from the input voltage to the output. In this device the input is the number of transfers for a section. Now non-linearity occurs because the size of the transferred charge packet changes with the gain-error voltage between the opamp inputs. This error voltage is again proportional to the number of previous charge transfers. The non-linear transfer is mainly generated in the first section while the error magnitude in the other sections is reduced by 2^{-p}, because the number of transfers in these sections has been made constant:

$$V_{out} \approx V_{LSB}(W_{LSB} + 2^p W_{MSB}) \frac{C_1 C_3}{C_2 C_4} \left(1 - \frac{W_{MSB} C_1}{2 A_{DC} C_2}\right) \tag{7.54}$$

With 80 dB opamp gain this effect is sufficiently reduced.

The settling of the charge transfer is important as it transforms clock jitter into output noise: the magnitude of the transferred charge from C_1 to C_2 is determined by the moment where the discharging of C_1 stops. With a limited unity gain of 6 MHz and the second pole of the opamp at 30 MHz the charge fraction which is not transferred is 5×10^{-4} after 100 ns. With an average of 80 transfers per conversion, the contribution to the noise of clock jitter to the total S/N ratio is in the order of −90 dB if the clock jitter is below 1 ns. The gain and settling requirements of the

Table 7.5 Measured performance of the algorithmic digital-to-analog converter

DC resolution (monotonicity)	15 bit
S/(N+THD)	74 dB
Dynamic range	87 dB
Sample frequency	44 kHz
Clock frequency	5.6 MHz
Power consumption (2.5 μm CMOS)	22 mW

opamp have been realized by means of a folded cascode configuration followed by a Miller stage. The input stage and the current source transistors contribute to the noise, and have to be designed carefully.

The results of the implementation are summarized in Table 7.5. Measurements have been performed with an external sample-and-hold circuit. The DC measurements show the inherent monotonicity of the 15-bit digital-to-analog converter. The noise and distortion figures of Table 7.5 include the sample-and-hold contributions.

The main drawback of this approach is that after the converter an additional sample-and-hold has to be used in order to create a time-continuous signal.

7.5 Digital-to-Analog Conversion in the Time Domain

7.5.1 1-Bit Digital-to-Analog Converter

Increasing the resolution of a digital-to-analog converter improves the accuracy with which the signal is reconstructed. Unfortunately this does not remove problems with linearity and distortion. In a multi-bit resolution converter, the transfer curve is made up of many physical levels, representing the relevant digital codes. Perfect linearity requires that all these physical levels form a perfect transfer curve. Even with just three physical levels it is impossible to construct an ideal transfer curve as three levels enclose two physical steps that are never perfectly identical (Fig. 7.52). Although it is counterintuitive the 1-bit digital-to-analog converter is the solution to the linearity problem in a digital-to-analog converter. In the reconstruction there are only two levels involved, and these form an ideal transfer curve. Variations in the levels affect the gain, not the linearity. Although the gain of such a conversion is less well defined, this idea finds its application in time-domain and pipeline converters.

7.5.2 Time-Domain Signals

Besides subdivision in the voltage, current or charge domain, digital-to-analog conversion can also be realized by means of time division. The signal information is contained in the succession of switching moments, see Fig. 7.53. Time-domain conversion often uses 1-bit digital-to-analog conversion.

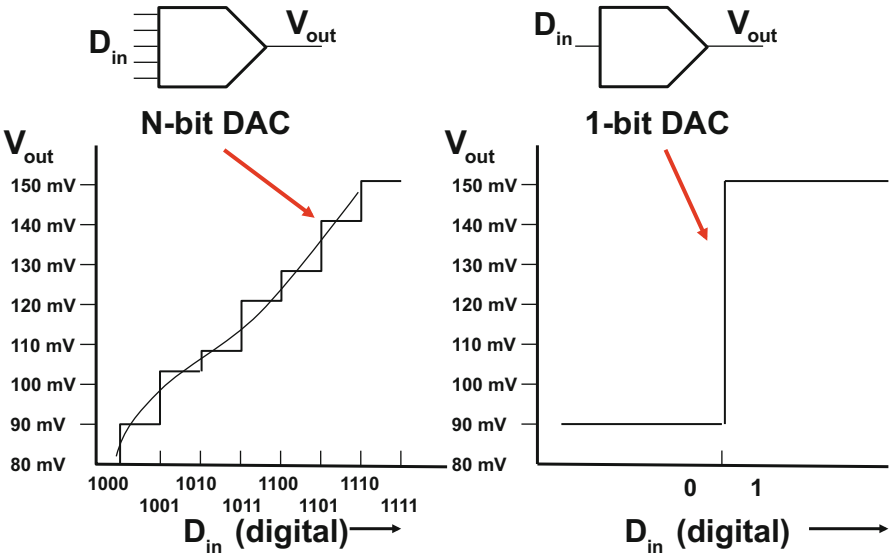


Fig. 7.52 An N-bit and a 1-bit digital-to-analog converter differ fundamentally on distortion

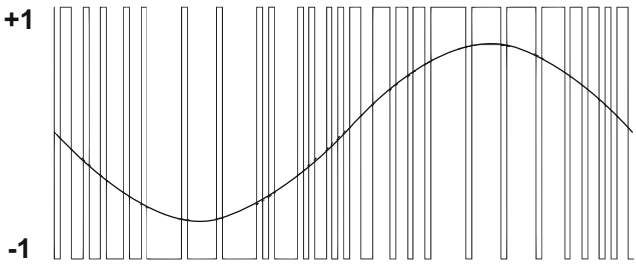


Fig. 7.53 The low-frequency content of a pulse sequence contains the signal

One of the first techniques to digitize information was based on the succession of pulses [47] and was originally called pulse code modulation (PCM). This specific implementation is today more often referred to as pulse width modulation. In pulse density modulation (PDM) the signal is composed of varying numbers of equal pulses. These techniques employ a processing rate that is larger than the Nyquist sample rate. The ratio between both is called the oversample ratio (OSR).

All time division schemes have in common that the output switches between a few (two) levels of a physical medium (voltage or current). Any DC-variation on these levels will manifest itself as a gain factor and will not affect the conversion quality. The linearity problems due to component inaccuracies are circumvented. Pulse width modulation or pulse density modulation only uses two levels to create a fast switching pulse sequence. The high and low time of the digital pulse train is ordered in such a way that the low-frequency component of the signal is correctly

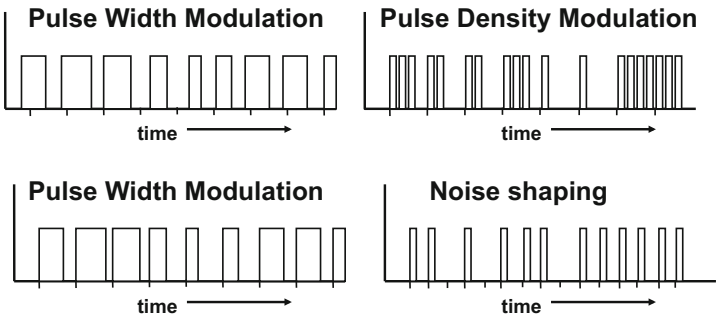


Fig. 7.54 Several forms of time-domain digital-to-analog conversion

represented. In a simple form a pulse-width or pulse density modulation (PWM and PDM) can be converted into the analog domain by filtering the pulses in a suitable low-pass filter.

Figure 7.54 shows four differently coded sequences. On the left side are the pulse width modulation formats. The amplitude of the signal is proportional to the width of the pulse. These pulses can be synchronized to the sample rate or free running. PWM signals are not quantized in amplitude: the pulse width is proportional to the original amplitude. On the right-hand side pulse-density modulated signals are shown. In PDM the amplitude is quantized. Pulse density avoids issues with distortion due to asymmetries in falling and rising edges by using just one type of pulse. Sigma-delta modulation and noise shaping are forms of pulse density modulation.

Whatever form of time-domain conversion is used, it is important to realize that a large portion of the available energy in the time-domain representation is unwanted:

$$\begin{aligned} \text{Energy in PWM signal} \quad &\propto (+/-A)^2 \quad = A^2 \\ \text{Energy in maximum sine wave} &\propto \int (A\sin(\omega t))^2 dt = A^2/2 \end{aligned}$$

Even when the time-domain signal is representing a full-scale sine wave half of its energy consists of spurious components and needs to be removed.[12] These unwanted signal components are normally located at higher frequencies, however if these frequency components are applied to non-linear elements, down mixing of harmonics of the fundamental signal can occur.

The accuracy of generating time moments is limited as well. Jitter affects the actual position in time of the transitions. As jitter normally is a signal independent process, the accuracy of the conversion improves at low signal frequencies.

[12]The ratio between signal power and the power of the harmonics of a perfect block wave is $1/(\pi^2/8-1) = 4.27$ or 6.31 dB.

Any misplacement of a switching edge is directly multiplied with the maximum signal swing. The effect of jitter on a pulse-density coded signal is therefore considerably larger than in the sampling of analog signals, compare Sect. 2.6.1. The maximum sinusoidal signal power that is contained by a pulse train switching between $+A$ and $-A$ is the power of $A\sin(\omega t)$. The time jitter is σ_{jit} which results in $2A\alpha\sigma_{jit}f_s$ after multiplication with the amplitude and normalization with respect to the sample rate.[13] $0 < \alpha < 1$ is an activity factor indicating the fraction of transitions compared to the sample rate. This gives a signal-to-noise ratio of:

$$\mathrm{SNR} = \frac{\frac{1}{T_a}\int_{t=0}^{T_a}(A\sin(\omega t))^2 dt}{\frac{1}{T_a}\int_{t=0}^{T_a}(2A\sigma_{jit}\alpha f_s)^2 dt} = \frac{1}{8(\alpha f_s\sigma_{jit})^2} = \left(\frac{1}{4\sqrt{2}\mathrm{OSR}\alpha f_{sig}\sigma_{jit}}\right)^2 \tag{7.55}$$

with $\mathrm{OSR} = f_s/2f_{sig} \gg 1$ as the oversample ratio. Comparison with the signal-to-noise ratio of sampled signals in Eq. 2.36:

$$\mathrm{SNR} = \left(\frac{1}{2\pi f_{sig}\sigma_{jit}}\right)^2 \tag{7.56}$$

shows that the impact of jitter on PWM signals is related to the sample rate. Therefore jitter in time-domain digital-to-analog conversion is an order of magnitude higher than in voltage or current-domain converters, see Fig. 7.55.

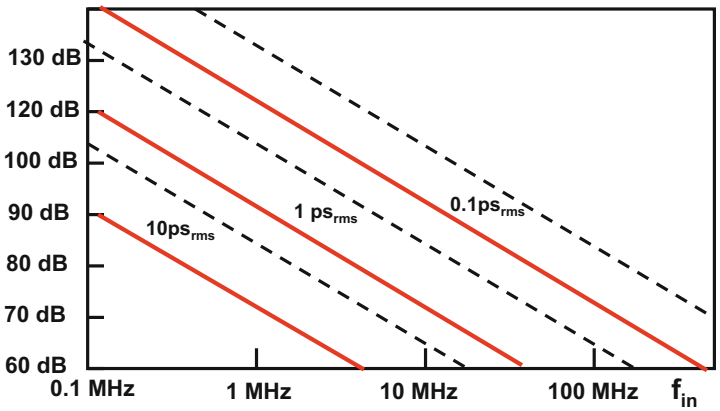


Fig. 7.55 Comparison of jitter in normal sampling (*dotted line*) and in one-bit signals where a moderate oversampling ratio of 5 is used

[13] Assuming only one edge is jittering.

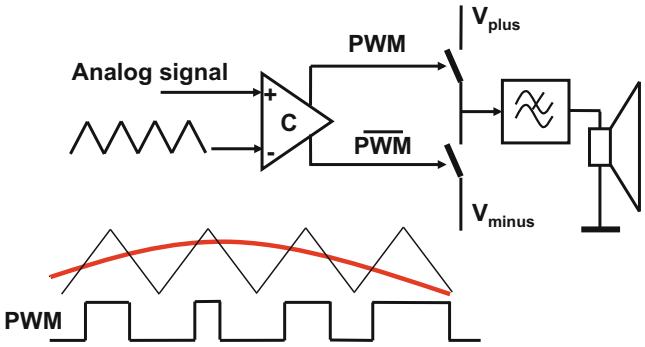


Fig. 7.56 Class-D amplifier with PWM modulation

7.5.3 *Class-D Amplifiers*

A well-known application of pulse-width modulation is in audio class-D amplifiers as shown in Fig. 7.56. The incoming analog signal is compared to a triangular signal (as can be generated from the integration of a block wave). At the crossings the PWM signal flips polarity. This PWM signal is used to drive the switches that connect to the positive and negative power. This power-PWM signal is low-pass filtered and applied to the loudspeaker. Potentially this method can result in 85–90 % efficient output stages. More advanced schemes are reported in, e.g., [168].

The above scheme illustrates a partial transition between analog, time-coded analog and analog again. Analog-to-digital and back, but without explicit time and amplitude discretization.

Example 7.17. What efficiency can be expected from a PWM class-D output stage for 4 Ω load impedance with 0.2 Ω resistance per switch, 20 nF switch gate capacitance, and 0.5 MHz clock frequency.

Solution. The major loss mechanisms in class-D are: IR- drops, R_{on} of the switch and switching losses of both switches: $CV_{DD}^2 f$. For the total expected efficiency this means:

$$\eta = \frac{P_{load}}{P_{load} + P_{res} + 2P_{cap}} = \frac{I^2 R_{load}}{I^2 R_{load} + I^2 R_{on} + 2CV_{DD}^2 f}$$

substituting $V_{DD} = I(R_{load} + R_{on})$ allows to eliminate the current and an efficiency of 88 % is found.

7.6 Methods to Improve Accuracy

Small variations in parameters as discussed in the previous section lead to additive and multiplicative errors in analog circuits. In an amplifier additive errors can be regarded as a DC-offset, while multiplicative errors affect the overall gain.

The digital-to-analog converters in the previous paragraphs consist of a division mechanism generated by multiple copies of components (e.g., a resistor string or a set of current sources) and a selection mechanism that chooses the appropriate settings. The signal path differs for every new signal and errors will appear in a certain range of the signal, see Fig. 7.57. Some of these errors can be seen as addition of an error value. In more complex converters with subranges, like current-steering converters or successive approximation converters with bridge capacitors, gain differences may appear between multiple sections of a digital-to-analog converter. These errors generate complex and signal-dependent patterns that result in distortions, spurious tones, or noise-like behavior. With the design guidelines of Table 5.5 the systematic errors in components can largely be eliminated. And large devices reduce the random error to a level that allows a 10- to 12-bit accuracy. If more accuracy is needed, additional measures are required in the design.

The discussion in this section is limited to mitigation of additive errors. The impact and avoidance of multiplicative and timing errors is further discussed in Chap. 9.

In Fig. 7.58 the errors $\epsilon_{1,2,3,i}$ affect the different paths through the analog-to-digital or digital-to-analog converter. Four classes of mitigation of additive errors are shown.

The first method to mitigate additive errors was 50 years ago already in use for precision instrumentation. The chopping technique [169], Fig. 7.58 (top), first modulates the input signal to a higher frequency band after which this signal is processed by the converter (or any other form of signal processing). Any error located at DC or in a low-frequency band will not interfere with the signal.

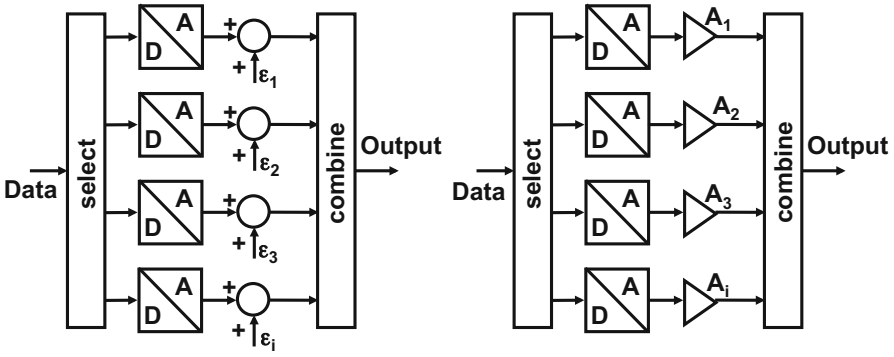


Fig. 7.57 Additive and multiplicative errors in digital-to-analog conversion

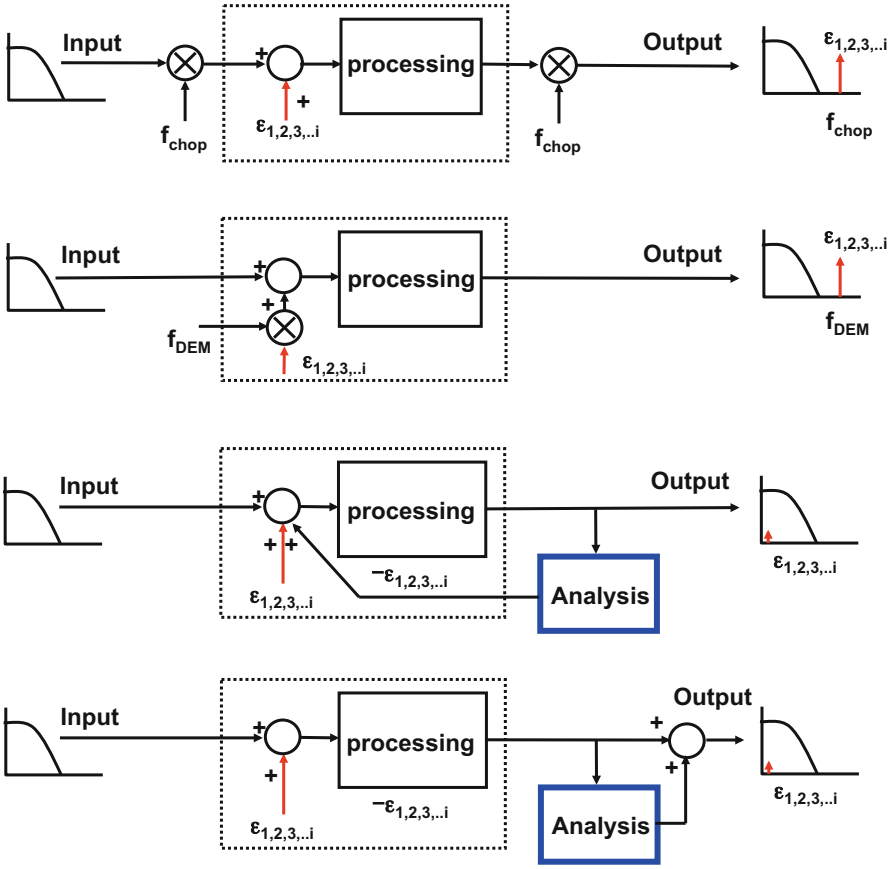


Fig. 7.58 Four methods to mitigate additive errors. *Top*: chopping, *second*: dynamic element matching, *third*: calibration with feedback, *bottom*: calibration with feed-forward

Now the processed and error signal are modulated back with the same frequency as used at the input. The signal is restored to its original frequency range and the additive error is modulated to the chopping frequency. This technique relies on good quality mixers/modulators and of course requires the conversion function to process the signal at that frequency. Although this technique is rarely used in a converter [170], it is often implicitly present in digital radio systems. In contrast to calibration the error is still present and its energy may come back into the signal path after distortion, see also Sect. 2.3.1.

The second principle addresses the error at its source, just as in calibration. The error is not removed but moved to a less critical frequency range. Examples of this technique are dynamic element matching and data-weighted averaging, Sects. 7.6.2 and 7.6.3.

The two calibration principles (Fig. 7.58 (bottom)) use a feedback or feed-forward path to compensate the errors at the node or close to the node where they originate. The feedback is generated by means of a form of error measurement. The feedback loop can take many forms. For example, a single measurement can be done during the production test and the feedback consists of a laser trimming some resistor value. Another option is a form of on-chip measurement, which runs continuously during the life time of the device, see, e.g., Fig. 8.104. The advantage of calibration is that once in place the remaining circuit acts as an error-free circuit. The error is removed, not only moved. The electronic feedback path in such a system is continuously present and may introduce noise into the signal path. An example of calibration in a matrix of current sources is given in Sect. 7.6.1.

An alternative form of calibration measures the output signal's error and subtracts it. A nice example is the cascaded sigma-delta modulator in Sect. 10.4.6.

In contrast to DEM and DWA calibration is characterized by measuring the error or comparing two values. Calibration is therefore limited by the accuracy of the measurement or comparison.[14] Off-chip calibration allows ppm level accuracies and other mechanisms such as drift limit the performance. On-chip calibration is more tedious and often limited to 14–15 bit level. DEM and DWA avoid this issue.

7.6.1 Current Calibration

With the change from bipolar devices to CMOS techniques with easy capacitive storage, new forms of calibration became possible. Figure 7.59 shows a unary current array with a current divider as in Fig. 7.30. The unary row of current sources is extended by one, from 64 to 65. This allows to calibrate every clock cycle one of the current sources of the unary array. In Fig. 7.60 two NMOS transistor form the calibrated current source: M_1 supplies the majority of current as a traditional current source. M_{1a} is used to calibrate and to mitigate the inherent mismatch in these sources. During the calibration the spare current source is connected as a sort of track-and-hold circuit by closing the switch between gate and drain. Feeding this arrangement with a reference current will force the gate voltage of M_{1a} to settle at the level needed to balance this current. After opening the switches the gate capacitor will store a gate voltage that generates a current in M_1 and M_{1a} equal to the reference current. This current source is ready for use. Another current source is taken out of the array and will be calibrated. In this way the calibration mechanism rotates through the array tuning every 65 cycles all current sources. The unary array of calibrated current sources in the digital-to-analog converter of Fig. 7.59 is completed by a 10-bit passive current divider to yield a 16-bit digital-to-analog converter. The main problem in this arrangement is the sampling of noise in the calibration cycle. Unwanted low-frequency noise as $1/f$ noise is suppressed

[14] From a discussion with Lucien Breems for the AACD 2016 pannel.

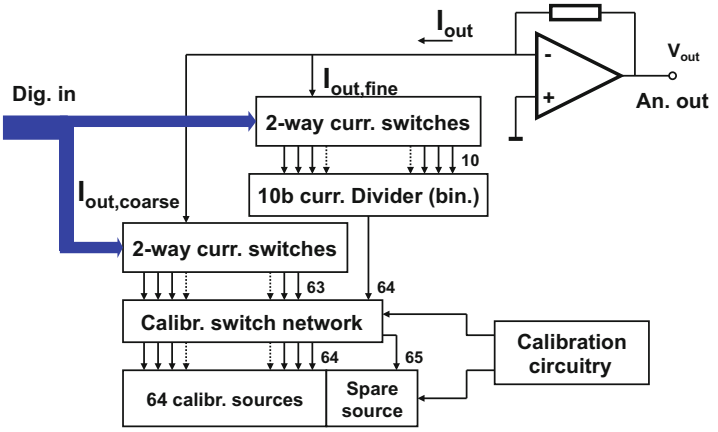


Fig. 7.59 An example of a digital-to-analog converter applying current calibration [134]. This design was produced as TDA1545

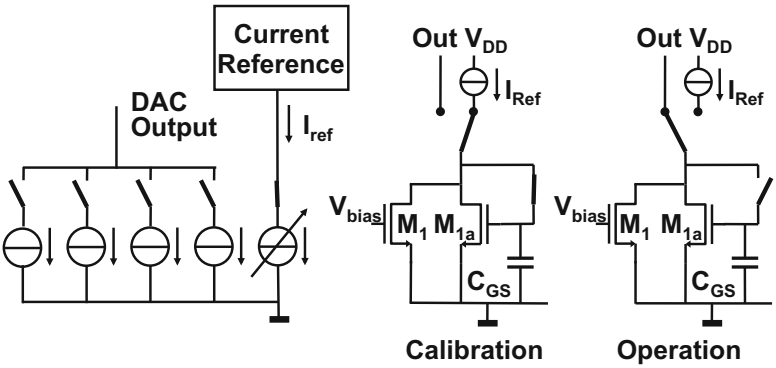


Fig. 7.60 The current calibration technique [134]

up to half of the calibration frequency, but the remaining contributions influence the current considerably. This is mainly due to the relatively low calibration frequency which is experienced by the individual current sources. For that reason most of the current is generated by the standard current source M_1 and only a few percent is calibrated via M_{1a}.

7.6.2 Dynamic Element Matching

The dynamic element technique swaps identically designed voltages and currents in the circuit topology. In Fig. 7.61 the two currents I_1 and I_2 have been created by dividing a main current. Although these currents are equally designed, they show

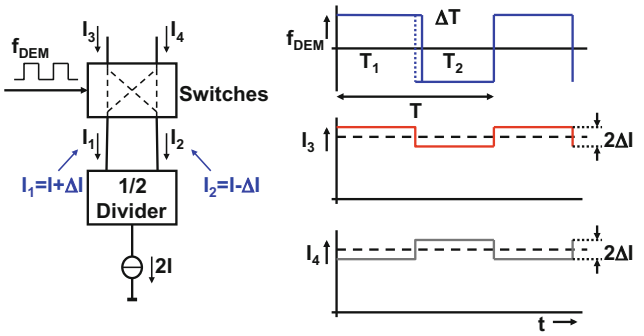


Fig. 7.61 Dynamic element matching [130]

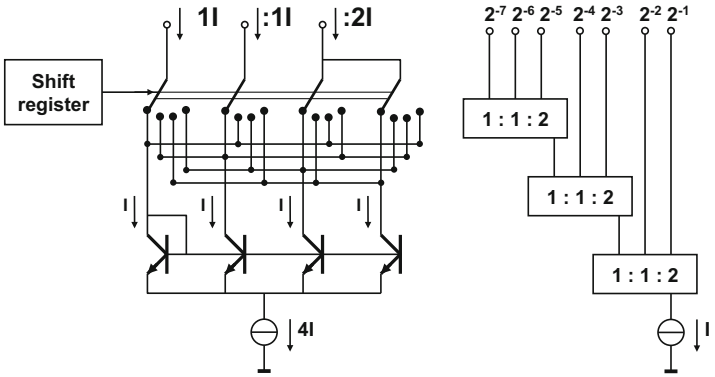


Fig. 7.62 An example of a digital-to-analog converters applying dynamic element matching [133]. This idea was the basis for the first Compact Disc audio digital-to-analog converter TDA1541

a small offset with respect to each other due to random variations. A switching box alternates the connection to the outputs. The resulting currents I_3 and I_4 are composed of both I_1 and I_2 and average out the difference over time. The resulting error in I_3 and I_4 is the product of the original inequality times the duty cycle of the switching pulse.

$$\frac{I_3 - I_4}{I_3 + I_4} = \frac{I_1 - I_2}{I_1 + I_2} \times \frac{t_1 - t_2}{t_1 + t_2} \tag{7.57}$$

This technique is known as "dynamic element matching" [130] and is applied in many types of converters. Figure 7.62 shows an extension of the technique to generate three currents in the ratio 1:1:2. Cascoding a number of these stages leads to a 16-bit current digital-to-analog converter. In this design variant stacking multiple D.E.M. circuits requires voltage head room. Therefore the dynamic element technique is today more used in order to tune single groups of devices. Note that the swapping frequency will not interfere with signal path: here only the

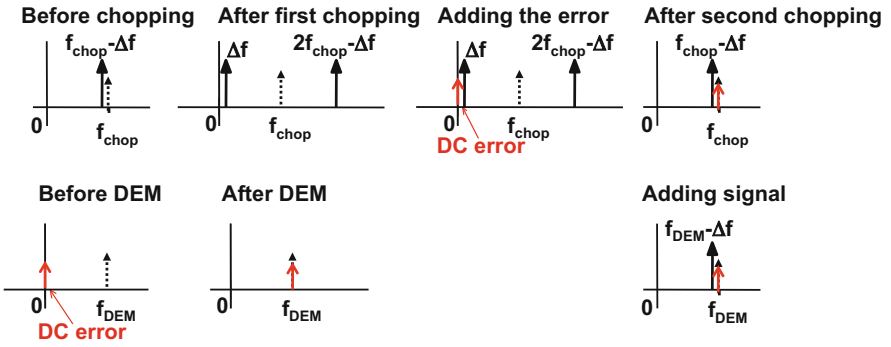


Fig. 7.63 Comparison of chopping (*above*) and dynamic element matching (*below*) of signals close to the DEM or chop frequency

exponentially weighted currents are formed that are the basis for a binary digital-to-analog converter. The switching for the actual conversion is performed after the D.E.M. circuit and spurious frequency components can be filtered out before the signal switching.

Example 7.18. What happens to an external input signal near to the chopping frequency? And what happens to an input signal close to the DEM frequency?

Solution. Figure 7.63 shows the chopping mechanism in comparison with the dynamic element matching. Chopping moves the signal frequency to a higher frequency before the error is introduced. After that the error and the up-modulated signal frequency are modulated again, causing the signal frequency to appear at its original position and the error at the chopping frequency.

The dynamic element mechanism directly up-modulates the error, before it is added into the signal chain.

7.6.3 Data-Weighted Averaging

In a straight-forward unary digital-to-analog converter, the taps of a resistor string, the capacitors in a capacitive array or the currents in a current source array are selected starting with the first, second etc. element until the sample magnitude is reached. This process is repeated for every new sample. The signal in the digital-to-analog converter depicted in Fig. 7.64 (left), is built up of perfectly equal current sources except for one. The somewhat larger third source creates a positive DNL error at a fixed position in the signal. All samples larger than this value will be shifted by $DNL \times I_{LSB}$ and create harmonic distortion, as is shown in Fig. 7.6.

The previously discussed dynamic element matching technique swaps the various components in a pre-determined order and under control of an independent

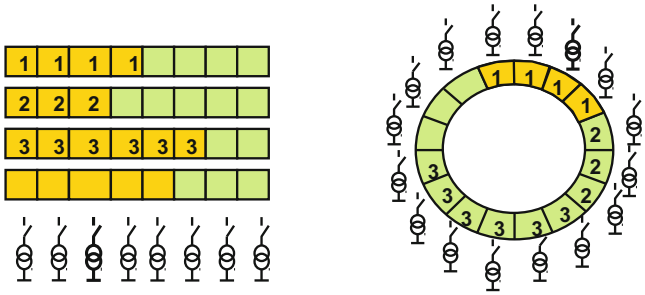


Fig. 7.64 Data-weighted averaging [172, 173]

frequency. The DEM idea is to transform a fixed DC-error at a fixed position in an array of resistors, capacitors or currents into a modulated component at a high frequency. This spur can either be filtered out or moved to a frequency band where it poses no problem.

Another attempt to solve the problem of a fixed error in the array is to change the starting point for the signal in the array. For example, the next sample starts at an offset of one position. This offset is increased for every next sample. The offset can even be chosen randomly [171]. These techniques transform the error at the third position in this example into a sort of noise as the correlation with a fixed position in the signal amplitude is avoided. Mathematically this principle can be modeled by assuming that the current sources $I_{LSB,i}, i = 0, \ldots \infty$ form an infinitely long line of current sources each with an independent random error σ_I. A signal sample formed with p of these current sources is disturbed with an error $\sigma_I \sqrt{p}/I_{LSB}$. If the average value of p is around mid range $E(p) = 2^N/2$, the long-term average noise per sample is $\sigma_I \sqrt{2^{N-1}}/I_{LSB}$. In the frequency domain this results in a noisy spectrum that becomes flatter with an increasing number of independent current sources.

This idea is taken one step further in the data-weighted average (DWA) principle [172–175].[15] The current sources are now functionally arranged in a circle, see Fig. 7.64 (right). Every next sample starts after the last current source of the previous sample.

An intuitive analysis starts with the assumption that there is only one deviating current source I_p with an error Δ_{Ip}, see Fig. 7.65. The error signal due to Δ_{Ip} will appear in the output at a rate depending on the number of times the current source I_p is requested. This results in an unwanted spurious tone at f_{Ip} where the frequency depends on the level of the output signal. For example, for a DC-signal level at one-quarter of the full-scale, the error will appear once in every four sample pulses, so the error component will appear around $f_s/4$. With an average signal around

[15] Two sources are available as the originator of DWA: Michael Story [174] and Maloberti [175].

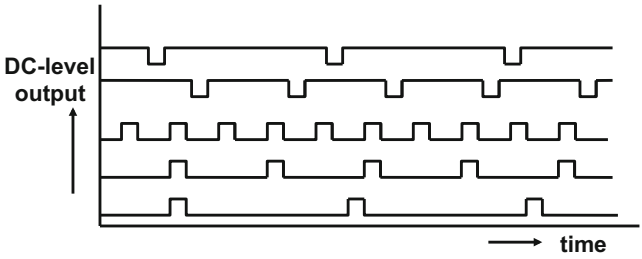


Fig. 7.65 A single mismatching current source will appear at a low frequency if the average signal value is close to the extremes of the range

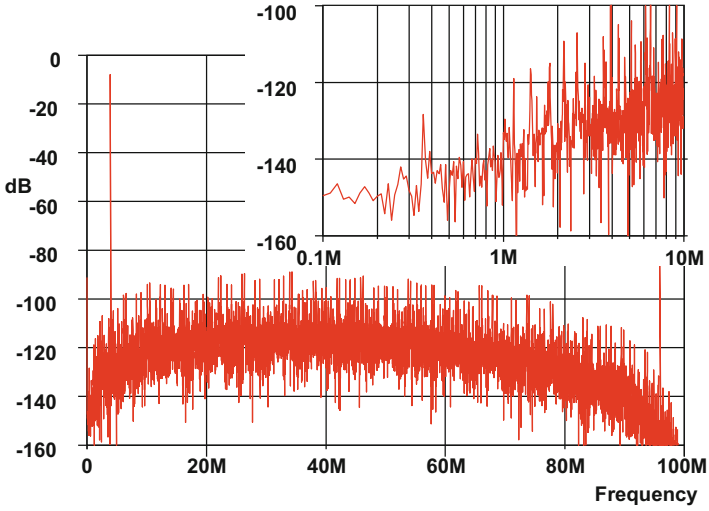


Fig. 7.66 Spectrum of a data-weighted averaging digital-to-analog converter. 256 current sources are simulated with a normally distributed mismatch of $\sigma_I/I_{LSB} = 1\,\%$. The signal frequency is 3.97 MHz at 100 Ms/s. In the insert a small fraction of the range is visible. The spectrum contains a lot of tones

mid-range the error will produce tones around $f_s/2$ and generate maximum error power. The error frequency and the error power associated with it are proportional:

$$P_{Ip} \propto f_{Ip}\Delta_{Ip}^2 \tag{7.58}$$

For a single error the power of the spurious component is proportional to the frequency. In a spectrum this leads to a first-order behavior of the error tones versus frequency.

If there is no dominant error source, but a collection of errors, the spectrum of the random components will also show this first-order frequency shaping, as is seen in Fig. 7.66.

The formal proof [172] requires the introduction of some stochastic mathematics beyond the scope of this book. The following reasoning summarizes the line of thought without pretending to present a solid mathematical proof.

If an array is composed of 2^N current sources with random varying values $I_{LSB,i}, i = 0, \ldots 2^N - 1$, the cyclic operation causes that in the long term every current source is used an equal amount of times. Therefore it is useful to specify the average or mean value of every current source by means of the expectation value:

$$m_I = E(I_{LSB,i}) = \frac{1}{2^N} \sum_{i=0}^{i=2^N-1} I_{LSB,i} \tag{7.59}$$

This formulation modifies the ideal amplitude of the actual current source I_{LSB} into the mean value m_I of a group of 2^N current sources. For an infinite number of current sources m_I converges to I_{LSB}, but for a specific converter with a limited number of sources $m_I \neq I_{LSB}$. This algorithm may (slightly) change the amplitude of the signal compared to the ideal case, in practice the result will be a minor change in gain. In fact every converter redefines its own I_{LSB} and slow signals (where all current sources are used in roughly equal numbers) will not be degraded.

A sample at time $t = nT$ contains k current sources and its output value will be

$$I(nT) = \sum_{i=0}^{i=k-1} I_{LSB,i} = \sum_{i=0}^{i=k-1} m_I + \sum_{i=0}^{i=k-1} (I_{LSB,i} - m_I) = k \times m_I + \sum_{i=0}^{i=k-1} (I_{LSB,i} - m_I) \tag{7.60}$$

In the result the term $k \times m_I$ is the desired signal value of the digital-to-analog conversion and the last term in the summation represents the unwanted or noisy part of the array of current sources. If k is a fixed odd value[16] (a DC signal), then there is an addition to every sample specified by the last term in Eq. 7.60. For DC signals this term converges to zero for an infinite number of samples.

The second term of Eq. 7.60 defines the error for a sample at time $t = nT$. As the total number of sources is bounded to 2^N, the error sum of the remaining sources is given by

$$\sum_{i=k}^{i=2^N-1} (I_{LSB,i} - m_I) = \sum_{i=0}^{i=2^N-1} (I_{LSB,i} - m_I) - \sum_{i=0}^{i=k-1} (I_{LSB,i} - m_I) = - \sum_{i=0}^{i=k-1} (I_{LSB,i} - m_I) \tag{7.61}$$

Given the circular structure it is obvious that the remaining error has the magnitude of the error at time $t = nT$ with an opposite sign. This is true because m_I acts as the effective LSB value. In order to construct the next sample(s) the circular arrangement uses these remaining sources with their errors. So this error will show up in the next (few) samples, depending on the values of the succeeding signal

[16] Odd value: to stay away from idle patterns at this stage of the explanation.

samples. If the delay of the remaining error term with respect to the originating error term is estimated as αT_s, with α is close to unity, the error transfer function of the error becomes

$$H_{\text{error}}(z) = \sum_{i=k}^{i=2^N-1} (I_{LSB,i} - m_I)(1 - z^{-\alpha}) \tag{7.62}$$

where the term $1 - z^{-\alpha}$ corresponds to a first order high-pass filter. This result for a single sample must be generalized for all α, k and extended to all errors selected during the conversion of a complex signal. The overall frequency filtering of the error spectrum is then in first order:

$$|H_{\text{error}}(\omega)| \propto \sin(\omega \alpha T_s/2) \tag{7.63}$$

Along the same lines, higher order suppression schemes ($1 - z^{-2}$) can be designed [173]. Figure 7.66 shows an example of a first-order data-weighted averaging operation. The harmonic distortion products (in Figure 7.6 at −60 dB) have been considerably reduced due to the randomizing effect of the selection. Secondly the DWA algorithm has cleaned the spectrum around DC, see the insert. The noise level at higher frequencies $f_s/2$ has increased as is inevitable in "noise-shaping": the error power is moved, not removed. The consequence is that DWA must be applied in oversampling applications: the sample rate must largely exceed the signal bandwidth in order to allow a good bit of frequency band for the excess error power. Other switching sequences of the DWA algorithm allow to choose the frequency of maximum suppression at arbitrary points in the spectrum [161]. Miller and Petrie [176] shows a high speed implementation of the digital decoding.

Unfortunately the first-order DWA algorithm will not completely randomize the error signal. Figure 7.65 indicated already that a large error generates tones in the spectrum. The possibility that a fair number of samples of a DWA digital-to-analog converter produces tones is likely.[17] As a result often the DWA algorithm is extended with additional randomizing algorithms like the addition of dither, see Sect. 4.3.4 and [177].

Some resemblance is present between the DEM and DWA methods: swapping between 2^N components in DEM is like a DWA without the randomizing effect of the signal. A DEM with a fixed swapping frequency generates an error spectrum with tones at the multiples of the swapping frequency. Some designs use a pseudo-random swapping sequence to spread out the error energy over the spectrum.

[17] In Chap. 10 "idle patterns" are discussed. The patterns in data-weighted averaging bear a lot of resemblance but come from a completely different origin.

Exercises

7.1. If the 10-bit digital-to-analog converter of Fig. 7.20 is constructed from a 4-bit coarse ladder with 16 resistors of 300 Ω and 16 6-bit fine ladder sections with 64 resistors of 50 Ω each, calculate the maximum resistance in this ladder to the grounded reference terminals.

7.2. In a resistor string digital-to-analog converter all resistors are 10 % larger. What is the change in INL? What other changes must be expected?

7.3. In a resistor string digital-to-analog converter all even numbered resistors are 2 % larger, while the odd numbered are on spec. What is the change in INL and in THD? What other changes must be expected?

7.4. Show that a differential read-out of a ladder with a second order gradient as in Fig. 7.15 (left) shows third-order distortion.

7.5. In a resistor string digital-to-analog converter all resistors are randomly 1 % larger or smaller. What is the change in INL, THD, and SNR? What other changes must be expected?

7.6. The divider in the lower schematic of Fig. 7.29 has a redundant transistor "1T". What are the consequences if this transistor is removed?

7.7. Compare the binary sections of Fig. 7.29 (upper and lower). Make an estimate of the DNL difference for an equal number of bits in the binary sections.

7.8. A 5-bit fine resistor string is directly via switches connected to a 5-bit coarse resistor string. As a consequence current from the coarse string will run via the fine resistor string. The resistors of the coarse string are 100 Ω each. What should be the value of the fine resistors if the maximum DNL due to the resistive loading of the coarse string must be less than 0.5 LSB?

7.9. Connect in the previous example a current source to both ends of the fine ladder. Does this resolve the DNL problem? Is this solution free of other DNL problems?

7.10. An 8-bit unary current-steering digital-to-analog converter is driving a 1 kΩ load. Each current source has a parallel impedance of 5 MΩ and 0.2 pF. Sketch the resulting distortion behavior over frequency. Now the unused current is fed in a second 1 kΩ load allowing differential operation. There is a mismatch of 1 % between both load resistors. Sketch the distortion.

7.11. An 8-bit digital-to-analog converter based on a resistor string, suffers from a linear gradient of 1 % over the entire structure. What is the INL, DNL, and THD?

7.12. Replace in an R-2R ladder the resistors by MOS transistors. Under what circumstances can this MOS-2MOS ladder be used as a digital-to-analog converter? Estimate in a 0.18 μm technology the performance for a 0.5 V reference voltage.

7.13. In a current-steering digital-to-analog converter the current sources suffer from 5 % random mismatch. What DNL can be achieved for a 12-bit unary architecture? How many bits can be implemented in binary format is a DNL ≤ 0.5 LSB must be achieved for 99.7 % of the samples.

7.14. The current sources of a current-steering digital-to-analog converter must achieve a random mismatch of $\sigma_I/I < 1$ %. What are the W, L values for a CMOS65 transistor if $V_{GS} - V_T < 0.3$ V. Use the technological data from Table 4

7.15. Due to mismatches the INL pattern of an 8-bit binary architecture digital-to-analog converter shows steps of 1 LSB at 1/8, 2/8,... 7/8 of the scale. Calculate the distortion.

7.16. How much distortion is caused by a gradient of 2 %. This gradient means that at one end of a structure (e.g., a resistor string) the elements are 2 % larger than at the other end. It does not mean that every next element is 2 % larger.

7.17. A 12-bit current steering digital-to-analog converter is constructed of N_{uni} unary bits (=$2^{N_{uni}}$ current sources) and a $12 - N_{uni}$ bit binary section. Calculate the allowable capacitive load of an individual current source for a HD2 $= -60$ dB at $f_{sig} = 10$ MHz for various values of N_{uni}.

7.18. What unary–binary division in Example 7.10 results in a HD2 $= -60$ dB for a signal with $f_{sig} = 2$ MHz.

7.19. Due to unpredictable wiring patterns, the top-plate connections of the capacitors in Figs. 7.44 and 7.48 can experience a 1 % additional parasitic capacitance to ground. What will be the consequence for the achievable resolution?

7.20. A spurious level of -80 dB is required for a digital-to-analog converter. What is the maximum gradient that can be tolerated and what is the maximum random mismatch.

7.21. A digital-to-analog converter uses an output buffer. What is the maximum frequency sine wave of 1 $V_{peak-peak}$ that can be delivered to a 3 pF capacitor if the output current is limited to 100 µA? What changes if a 10 kΩ resistor is connected parallel to the capacitor.

7.22. A data-weighted averaging algorithm is used to eliminate conversion errors in a bandwidth located around $f_s/8$. Construct a sample sequence that will reduce the errors in that bandwidth, use [161].

7.23. Modify Example 7.11 by expanding the approximation to: $1/(1-a) \approx (1 + a + a^2) \quad |a| \ll 1$. Show that an odd distortion component remains.

Chapter 8
Nyquist Analog-to-Digital Conversion

The analog-to-digital converter compares the input signal to a value derived from a reference by means of a digital-to-analog converter, Fig. 8.1. The basic functions of an analog-to-digital converter are the time and amplitude quantization.

From the analysis of these processes in Chaps. 2, 4, and 7 some crucial limits for the design of an analog-to-digital converter have been identified:

- The thermal noise energy is given by: kT/C
- The jitter signal-to-noise power ratio is given by: $1/(\omega_{signal}\sigma_{jitter})^2$
- The quantization error is: $V_{LSB}^2/12$
- The mismatch between components is: $A_P/\sqrt{WL}$

The circuit ingredients that implement the sampling and quantization functions are the sample- or track-and-hold circuit where the sampling takes place, the digital-to-analog converter and a level-comparison mechanism. The comparator circuit is the location where the signal changes from its pre-processed analog form into a digital decision: "where nature turns into bits".[1] After the comparator a digital circuit processes the decisions into a usable digital signal representation.

Various subdivisions of analog-to-digital converters can be made. Based on the timing of the conversion four categories can be identified, see Fig. 8.2:

- The "flash" converters or "parallel search" converters in Sect. 8.2 use one moment in time for the conversion. The input signal and all the required reference levels must be present at that time moment. From a topological point of view no explicit T&H or S&H circuit on the analog side is required. The latches in the comparators form a sort of digital hold structure. Flash converters are very fast converters, however, the requirement to provide l 2^N reference levels and comparators leads to an exponential growth of the area and power of the converter. An elegant variant is the "folding" converter. Application that require

[1] Paraphrasing prof. Bram Nauta.

M. Pelgrom, *Analog-to-Digital Conversion*, DOI 10.1007/978-3-319-44971-5_8

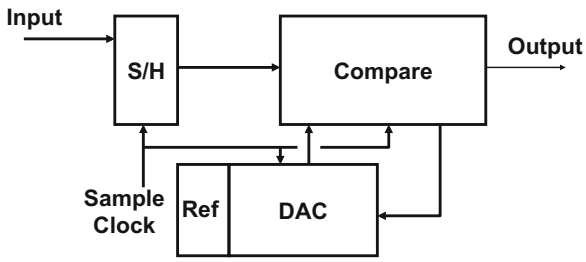


Fig. 8.1 Components in analog-to-digital conversion

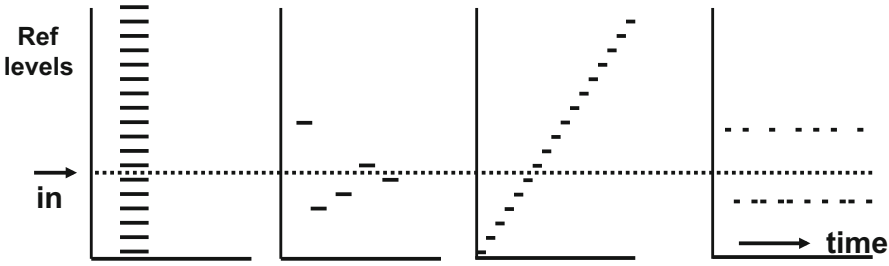


Fig. 8.2 Classification of analog to digital conversion principles. In the vertical dimension the available levels are depicted. *Left*: in "parallel search" conversion all levels are provided. The "sequential search" algorithm chooses the next level based on the information of the previous comparisons. In "linear search" each level is available at consecutive clock periods. *Right*: the "oversampled conversion" switches between a few levels

the highest speed and modest accuracy are served by these converters. Moreover, the flash converter is a basic ingredient in many other conversion topologies.

- The second category is the "sequential search" converter. This class builds up the conversion by choosing at every new clock strobe a new set of reference levels based on the information processed up to that moment, Sects. 8.3–8.6. For each step a suitable resolution can be chosen, mostly based on a power of base 2: 2^1, 2^2, 2^3 etc. A fundamental choice is to use the same hardware for all processing steps of one sample, or to use dedicated hardware for each next resolution step. In the last approach the total processing time for a sample is still the same, but pipelining allows multiple samples to be processed. Principles in this category are: successive approximation conversion, pipeline conversion, multi-step conversion. The combination of high accuracy and rather high speed makes these converters suitable for many industrial and communication applications.
- On the opposite side of the spectrum is the "linear search" converter, Sect. 8.8. All potential conversion levels are in increasing or decreasing order generated and compared to the input signal. The result is an extremely slow conversion, built with a minimum amount of hardware and tolerant to many forms of component variation. An example is the dual-slope converter. The robustness of these converters makes them popular for slow-speed harsh environments, such as sensor interfaces.

- The last category of conversion principles is mentioned here for completeness. The oversampled converters use mostly a few reference levels and the output switches frequently between those reference levels to create a time average approximation of the input. The accuracy comes from the time domain. These feedback-based delta-converters do not provide the conversion result at a determined point in time, but are accurate over a larger number of samples. The special techniques and analysis tools for these delta modulators are discussed in Chap. 10.

The first three categories are called "Nyquist-converters." These circuits convert a bandwidth[2] close to $f_s/2$ and often operate at the speed limit of the architecture and process. This chapter discusses first the decision making circuit: the comparator. After that the topologies of the flash, sub-ranging, successive-approximation and linear converters are examined.

Chapter 9 discusses time interleaving of multiple Nyquist converters and Chap. 10 the oversampling technique and sigma-delta conversion.

8.1 Comparator

Every analog-to-digital converter contains at least one comparator. Yet, there is little uniformity on comparator topologies. Many aspects need consideration when designing a comparator. These aspects may vary for every different application, for specific classes of signals, a technology, etc. No universal "one design fits all converters" comparator exists. There are as many comparator circuit variants as there are analog-to-digital converter designers. Still some general remarks on the design of the comparator can be made. Accuracy and speed are the global design parameters that have to be balanced versus the power consumption.

The fundamental task of a comparator is to amplify a big or small difference between its input terminals into a digital decision. In other words to extract the sign of the differential input signal. A number of requirements can be specified for a comparator:

- A large amplification is imperative as miliVolt or microVolt signal differences are translated in digital levels.
- A wide bandwidth for operating on high-frequency signals.
- Accuracy of various forms is required. In practical terms this means a low input offset, a low noise-figure, for 1/f noise as well as for thermal noise. Clocked comparators should not add uncertainty to the decision moment: a low timing jitter is required.
- A low power consumption, especially in analog-to-digital converters employing a lot of comparators.

[2]This bandwidth can start at DC, but also far above f_s using down-sampling in Sect. 2.3.2.

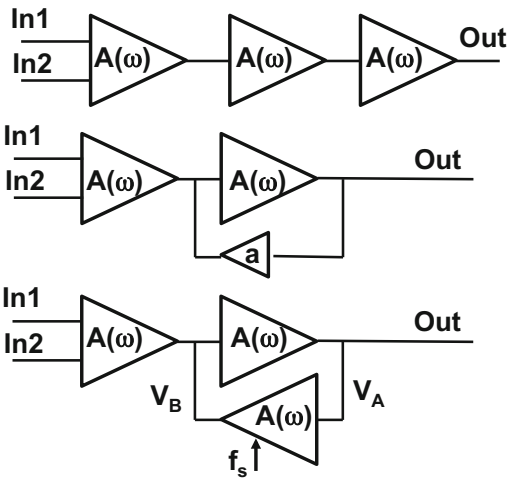


Fig. 8.3 Three basic comparator designs: straight-forward amplification, amplifier with hysteresis, and a latched or regenerative amplifier

- A wide input common mode voltage range, with a high common mode rejection.
- The previous comparator decision should not affect the following, no memory effect, also referred to as: inter-symbol interference.

Figure 8.3 shows three topologies for comparators. The straight-forward limiting amplifier (top) consists of a cascade of amplification stages to obtain as much gain as possible. For a given current consumption the unity-gain bandwidth (UGBW) per stage determines the overall speed of this chain of amplifiers. In amplifier theory the UGBW is that frequency where the gain equals unity.

If a single amplifier stage has a unity gain bandwidth ω_{UGBW} and an amplification A, and largely behaves as a first-order system up to ω_{UGBW}, the dominant pole is at $\tau = A/\omega_{UGBW}$. The response of a cascade of M comparators is

$$H(\omega) = \left(\frac{A}{1+s\tau}\right)^M = \left(\frac{A\omega_{UGBW}}{As+\omega_{UGBW}}\right)^M$$

An excitation with a step function results in

$$V_{out}(t) = A^M V_{in}(t=0)\left[1 - e^{-t/\tau}\sum_{i=1}^{M}\frac{(t/\tau)^{i-1}}{(i-1)!}\right] \tag{8.1}$$

For $t < \tau$ the cascade of comparators behaves as a cascade of integrators, therefore the response is proportional to $(t/\tau)^M$.

A second problem concerning the speed of this comparator is the recovery from the previous decision. Unless limiter circuits like diodes are used, the last stages of the amplifier will go into saturation if a large signal is applied to the input. This situation causes, e.g., the inversion charge of some MOS transistors in the circuit to be lost. The time required to resupply this charge leads to a delay during the next comparison.

Yet, the straight-forward limiting amplifier is popular as it requires no activation by a clock pulse. Often a string of inverters is used or some simple differential stages.

The second topology of Fig. 8.3 uses a small amplifier as a positive feedback element. As this amplifier is weaker than the feed-forward path it will only marginally contribute to the amplification. What it can do is add a threshold in the decision process. This so-called hysteresis is discussed in Sect. 8.1.7. Yet, in some circuit configurations this topology may arise, be-it unintentionally.

In the last topology of Fig. 8.3 the feedback path of the amplifier is of comparable strength to the forward path. The idea of this regenerative stage or latch is that the already build up difference in the forward stages feeds the positive feedback in order to reach a fast decision. This mode of amplification must be reset by a clock phase that clears the data after the decision, disables the feed forward, and allows pre-amplification.

The analysis of two positively fed back amplifiers or a latch is done in the Laplace domain. The nodes are labeled $v_A(t)$ and $v_B(t)$ or in the Laplace domain: $V_A(s)$ and $V_B(s)$. It will be assumed that there is an initial condition $v_B(t=0) \neq 0$.

$$V_A(s) = \frac{A}{1+s\tau}\left(V_B(s) - \frac{v_B(t=0)}{s}\right) = \left(\frac{A}{1+s\tau}\right)^2 V_A(s) - \frac{A}{1+s\tau}\frac{v_B(t=0)}{s}$$

where τ and A have the same meaning as above, but A is negative.

$$\begin{aligned} V_A(s) &= v_B(t=0)\frac{-A(1+s\tau)}{s((1+s\tau)^2 - A^2)} \\ &= v_B(t=0)\left(\frac{-A/(1-A^2)}{s} + \frac{A\tau/(2(1+A))}{1+s\tau+A} + \frac{A\tau/(2(1-A))}{1+s\tau-A}\right) \end{aligned}$$

The inverse Laplace transform gives

$$v_A(t) = v_B(t=0)\left(\frac{-A}{(1-A^2)} + \frac{A}{2(1+A)}e^{-(A+1)t/\tau} + \frac{A}{2(1-A)}e^{(A-1)t/\tau}\right) \tag{8.2}$$

After a few time constants have elapsed only the middle term in brackets is relevant:

$$v_A(t) \approx \frac{-v_B(t=0)}{2}e^{-(A+1)t/\tau} \approx \frac{-v_B(t=0)}{2}e^{\omega_{UGBW}t} \tag{8.3}$$

The nodes of the latch show an exponential increase determined by the start value and the unity-gain bandwidth. In order to reach a gain comparable to a cascade of M amplifiers with a gain A a time of $M\ln(A)/\omega_{UGBW}$ is needed. Ultimately the node voltage will be limited by the circuit and its power supplies. The exponential signal growth makes a latch a fast decision element in a regenerative comparator.

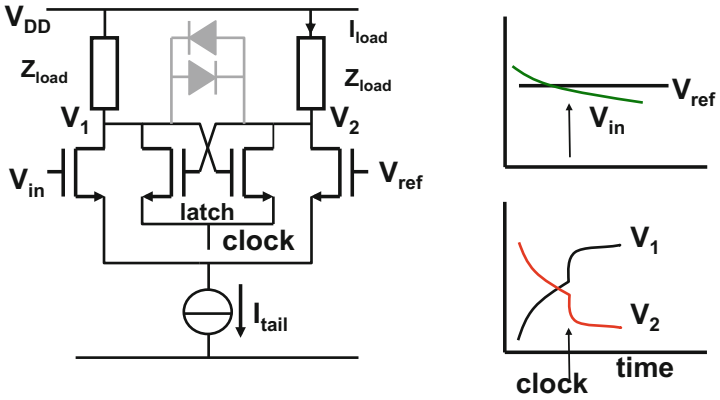


Fig. 8.4 A simple regenerative comparator circuit

8.1.1 Pre-amplification

Before the regenerative section of the comparator is activated, the signal is amplified. This amplification will reduce the influence of mismatch in the regenerative section and create a form of isolation between the regeneration process and the input sources.

Most comparators use an input differential pair. Here the difference is formed between the input value and the reference value. The differential pair allows to create some tolerance for the common mode level of the input signal. The differential current is applied to a load. After some amplification a positive feedback latch activated by a clock pulse will turn the latch on and amplify the small input voltage difference to a large signal. A basic example circuit is shown in Fig. 8.4. The gain of the comparator is determined by the transconductance of the input differential pair and the load: $g_{m,in} \times Z_{load}$. This gain must be sufficient to suppress mismatches, noise, etc. in the regenerative part of the circuit. On the other hand, there is no reason to boost the DC-gain to very high values. Too much gain will cause saturation, slewing and unnecessary kick-back. The achievable speed performance is limited by the unity-gain bandwidth, which in turn is determined by the parasitic capacitive load C_{load} on the internal nodes and the available current drive.

The bandwidth of the comparator must be analyzed both in small-signal mode and in large-signal mode. In small-signal mode the input transconductance and the capacitive load of the nodes V_1 and V_2 determine the bandwidth:

$$\omega = \frac{g_{m,input}}{C_{load}} \tag{8.4}$$

A large input transconductance is beneficial for a large small-signal bandwidth. Choosing a wide input transistor also helps in reducing the input referred mismatch,

but creates a larger capacitive load for the input terminal and the reference. Also the parasitic coupling between the drain voltages and the input terminals will become stronger increasing the kick-back.

In order to achieve the overall performance also the large-signal bandwidth of the circuit must be considered. Two main issues are crucial for the large-signal bandwidth:

- The fastest change of the signal that is amplified without distortion is limited by the slew-rate:

$$\begin{aligned} \text{slew-rate} = \frac{dV}{dt} &= \frac{I_{tail}}{C_{load}} \\ C_{load}\frac{dV_1(t)}{dt} &\leq I_{tail} \\ 2\pi f_{in} V_{1,max} C_{load} &\leq I_{tail} \end{aligned} \tag{8.5}$$

  where $V_{1,max}$ is the amplitude of an equivalent sine wave. If the charging current during the transient of the signal exceeds the tail current, the charging of the capacitors in the circuit will be limited and distortion can follow.
- A second large-signal effect in this comparator is saturation.[3] The large signals that drive a comparator will force the input transistors and the internal components in a saturated "on" or "off" regime. Internal nodes will be (dis-)charged to the power supply levels. Saturation of the input transistors creates significant currents in the gate connection, see Sect. 8.1.6. In order to revitalize the comparator all saturated components will have to be brought back into their linear operation regimes. This process requires current and the signal processing will experience a delay time. This will result in signal distortion. To prevent saturation effects, it is wise to limit the maximum swing on the internal nodes to 200–300 mV. In the comparator schematics of Fig. 8.4 two diodes are representing the signal swing limiting on the internal nodes from saturation. In normal CMOS technology these diodes are not available. Various other circuit techniques can be used to reduce the internal voltage swings (non-linear loads, cascode stages, etc.).

Figure 8.5 depicts a capacitive pre-amplifier. During a pre-charge phase, the capacitors are short-circuited and both internal nodes equal V_{DD}. When the switches open the current controlled by the differential pair will discharge the capacitors and the differential current causes divergent voltages. The gain due to the integration action is increasing with time. This circuit will not compare fluctuating input signals but operates only on stable signals from, e.g., a T&H circuit as in a successive approximation converter. The major advantage lies in the inherent filtering properties of the capacitors that limit the noise bandwidth [178, 179]. The switching operation on the capacitors creates kT/C noise. Next to that there is

[3] The term saturation is used to indicate that the circuit is far out of its operating point. Saturation of circuits has no relation with the operating regime of a transistor.

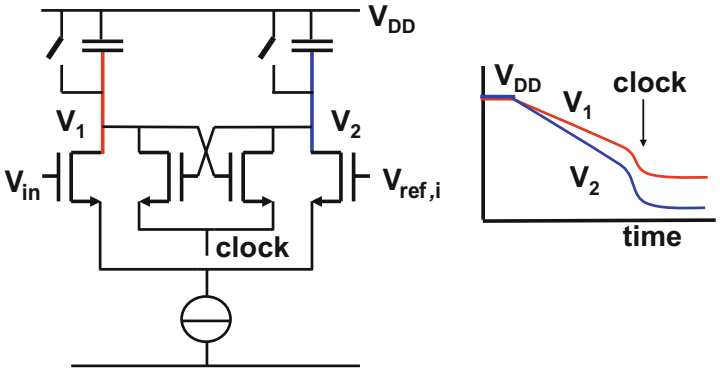


Fig. 8.5 A comparator circuit with capacitive amplification

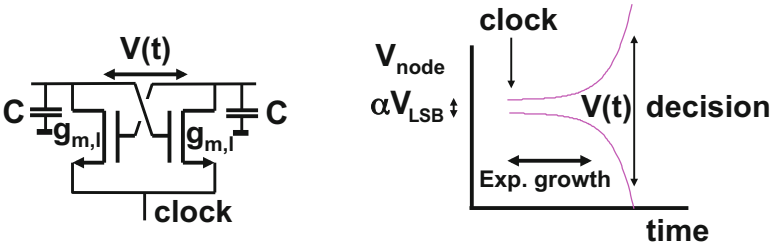


Fig. 8.6 The latch behaves as a linear feed forward amplifier for most of the decision time

the noise from the input pair. After the switches are released this noise is integrated on the capacitors and follows a random walk till after some time constants an equilibrium is reached.

8.1.2 Latch

A comparator in its regenerative phase is in the end a non-linear circuit: there is not a linear algebraic relation between input and output. However, in the early stages of the regeneration the feed forward circuit can be described with linear equations. Figure 8.6 (left) shows the basic ingredients: the transconductance of the latch transistors form a positive feedback system controlled by the time constant $\tau = C/g_{m,l}$. The starting point of the regeneration is the pre-amplified voltage αV_{LSB}, where $0 < \alpha < 1$ is a multiplier for an LSB size voltage. Now a simple Laplace analysis shows that

$$V(t) = \alpha V_{LSB} e^{+t/\tau} \tag{8.6}$$

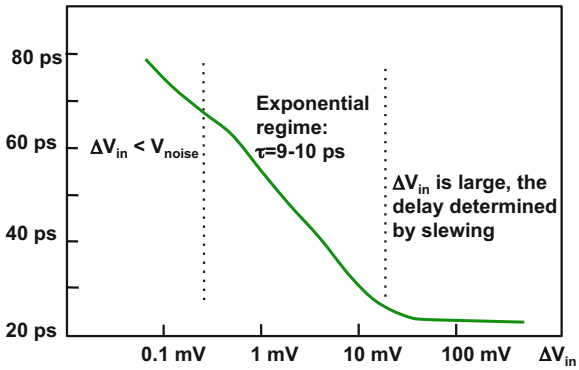


Fig. 8.7 The delay of a comparator decision as a function of the input voltage difference. Simulation in 65-nm CMOS. Courtesy: Jianghai He, MCCI

The exponential growth will continue until a physical barrier is reached, e.g. one of the nodes reaches a power supply, or some limiting mechanism takes action.

Another view is presented in Fig. 8.7. A popular comparator schematic has been simulated in a 65-nm technology. The overall delay time from the activation by the clock until an arbitrary large amplitude has been reached is shown as a function of the differential input voltage. At high input voltages, the delay by the regeneration is not relevant as other circuits cause a minimum delay. Below 10 mV the latch starts to create additional delay. As the input voltage is plotted on a logarithmic scale the delay shows a linear relation. Around 0.1 mV the circuit reaches the noise levels. This comparator with lay-out parasitics has a regeneration time constant of 9–10 ps.

8.1.3 Metastability and Bit-Error Rate

The regeneration delay in the comparator creates a fundamental problem in the form of metastability. Metastability is associated with any form of comparison and is also well known in, e.g., synchronization circuits. The crucial observation is that the time a comparator or latch needs to form a digital signal depends on the initial over-drive voltage. In other words: the input differential voltage determines the delay of the comparator.[4]

For very small overdrive voltages, there is a fraction α of the LSB size where the comparator has insufficient voltage difference on its nodes to reach a decision in the limited time T_s of a clock pulse. Now the comparator will not generate a clear "zero" or "one" output level after time T_s. With ambiguous signals the succeeding logical circuitry can generate large errors, e.g. if this digital signal is crucial for determining the MSB.

[4]This voltage-delay relation can be exploited: measuring the delay is used to quantify the input voltage and create more resolution.

For small signals the latch can be viewed as two single-pole amplifiers in a positive (regenerative) feedback mode. The voltages on the nodes of the latch develop exponentially in time with a time constant τ, see Eq. 8.6. This time constant is determined by the internal node capacitance and the transconductance of the latch transistors. Now the critical value of α below which there is ambiguity can be approximated by

$$\alpha \approx \frac{V_{latch}}{V_{LSB}} e^{-T_s/\tau} \quad \Rightarrow \quad \text{BER} \approx 2^N e^{-T_s/\tau} \tag{8.7}$$

where the ratio between the maximum latch swing V_{latch} and V_{LSB} is estimated as 2^N. The signal has a uniform probability arriving between 0 and V_{LSB}. If the range between 0 and αV_{LSB} causes errors, then the bit-error rate equals α.

The bit error rate is an important parameter in the design of fast converters with a high accuracy. In CMOS the time constant τ is formed by the parasitic and gate capacitances and the achievable transconductance.

A typical example with 8 bits assumes 5 fF total capacitance for 1 µm gate width, with 5 µA/V transconductance for the same gate width. A time period of $T_s = 20$ ns results in a BER of 10^{-7}. This BER can be improved to better than 10^{-8} by means of more current in the latch transistors. For converters with sample rates in excess of 100 Ms/s a BER of the order 10^{-8} means one error per second. Especially in industrial and medical equipment such an error rate can be unacceptable.

A method to get an impression of the bit error rate is to apply a slow sine wave to the device with an amplitude that guarantees that no more than one LSB change occurs at the digital output. Now the bit error rate can be estimated by recording output codes that differ more than one bit from the preceding code.

From a fundamental point of view[5] the BER cannot be avoided completely, however decreasing the time constant (by lower capacitance and higher transconductance), a BER of 10^{-13} is possible. Measures to improve the BER include: improving the latch speed with more current and smaller capacitances, additional gain stages or even a second latch, and a special decoding scheme avoiding large code errors due to a metastable state.

Example 8.1. Calculate the bit error rate of a latch in 0.18 µm CMOS with 0.5/0.18 µm NMOS transistors and 100 µA total current. The latch is used in a 0.5 Gs/s application.

Solution. The transconductance of the transistors is $g_m = \sqrt{2I(W/L)\beta_\square} = 0.3$ mA/V. The gate capacitance is $0.5 \times 0.18 \times 8.3$ fF $= 0.75$ fF. The succeeding stage will load latch with double or triple this amount. Some wiring and diffusion capacitance will result in a total load of around 10 fF, leading to a time constant

[5] Dr. Richard Schreier comment is that metastability is based on the idea that voltages can be scaled down continuously: mV, µV, nV,... However, is there a limit given by the quantum character on atomic level?

of 35 ps. The decision period is a 40 % fraction of the 2 ns sample rate. So the bit error rate is estimated at: $\approx 2^N e^{-T_s/\tau} = 2^N \times 1.2 \times 10^{-10}$. For a 7-bit resolution converter this would lead to BER $= 1.5 \times 10^{-8}$.

8.1.4 Accuracy

The required accuracy of a comparator depends on the required specification and the architecture of the analog-to-digital converter. In dual-slope, successive approximation and pipeline designs, the static random offset is a second-order effect and generates a (random) DC-shift of the entire signal. In a flash converter the input referred random offset is crucial because it will shift the individual reference levels and impacts both integral and differential non-linearities. Also in time-interleaved analog-to-digital converters, Chap. 9, the random offset returns as a spurious tone.

The input referred random offset σ_{Vin} must be designed to a value a factor 5 to 10 lower than the size of the LSB at the comparator as an initial target for e.g. flash converters. Fine-tuning of this requirement depends on, e.g., the tolerance for conversion errors on system level.

Accuracy has static and dynamic components. Just as in every analog circuit the static accuracy is (in a carefully laid-out circuit) determined by the random mismatch of a transistor pair, see Sect. 5.3. Of course the input differential pair is not the only contributor, also the mismatch of the load and the latch must be taken into account.

In the schematic diagram of Fig. 8.8 current differences caused by the input pair, the load and the latch all come together on the drains of the transistors. All contributions can be referred back to an equivalent input-referred random mismatch σ_{Vin}. With the help of Eq. 5.11, the input referred random offset is described as:

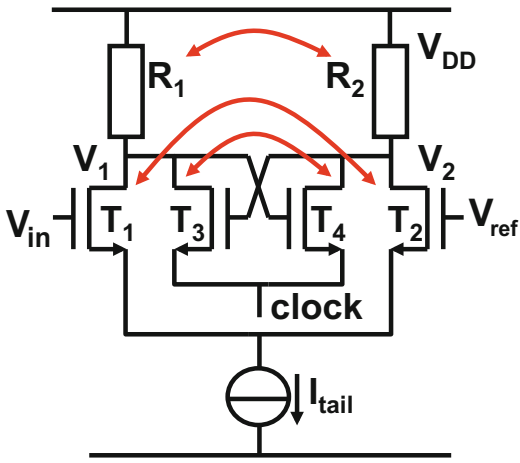


Fig. 8.8 Mismatch sources in a comparator

$$\sigma_{Vin}^2 = \left(\frac{\partial V_{in}}{\partial V_{T,12}}\right)^2 \sigma_{VT,12}^2 + \left(\frac{\partial V_{in}}{\partial V_{T,34}}\right)^2 \sigma_{VT,34}^2 + \left(\frac{\partial V_{in}}{\partial R_{1,2}}\right)^2 \sigma_R^2$$

$$\sigma_{Vin}^2 = \sigma_{VT,12}^2 + \frac{g_{m,34}^2}{g_{m,12}^2}\sigma_{VT,34}^2 + \frac{I_{load}^2}{g_{m,12}^2}\frac{\sigma_R^2}{R_{1,2}^2} \tag{8.8}$$

The equation assumes that the random mismatch can be calculated in a linearized model of the circuit. Unfortunately in this equation the transconductance of the latch does not behave in a linear fashion. In one extreme there is no current running in the latch and its contribution to the input-referred random offset is zero. Directly after the clock is activated and a current flows in the latch transistors, a step function on the nodes V_1 and V_2 will occur. If the gates are not equal or in the presence of mismatches due to differences in capacitive loading, different charges are drawn from the local nodes and will create random differences in these voltage steps.

If, on the other hand, the latch is constantly fed with a small tail current that is not sufficient to activate the latch, the dynamic errors can largely settle before the clock is activated. However the non-zero transconductance will contribute an additional component to the input-referred random offset.

8.1.5 Noise

Thermal noise and 1/f noise create additional accuracy limitations. Because these contributions are time-varying, also offset compensated analog-to-digital converter architectures suffer from this limitation. The contributions to the noise of the individual components are referred back to an equivalent input noise source in a similar manner as for mismatch. If the input transconductance dominates the noise, the effective value of the input-referred noise is:

$$v_{in,noise} = \sqrt{4kT\frac{BW}{g_m}} \tag{8.9}$$

with a bandwidth of 10 GHz and an input transconductance of 1 mA/V the value for the input referred rms noise is $v_{in,noise} = 400\,\mu V$. The thermal noise amplitude distribution creates a cumulative Gaussian probability distribution around the trip level, Fig. 8.9. In order not to exceed the accuracy specifications the bandwidth has to be reduced or more power has to be spent on improving the impedance levels.

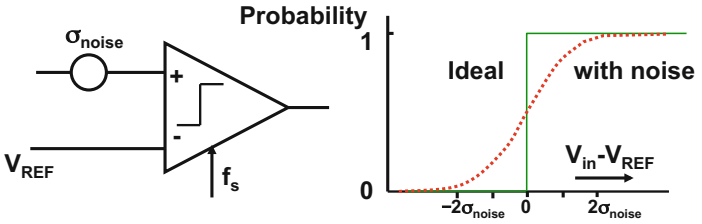


Fig. 8.9 Thermal noise will create a probabilistic behavior of the comparator decision (*dotted line*)

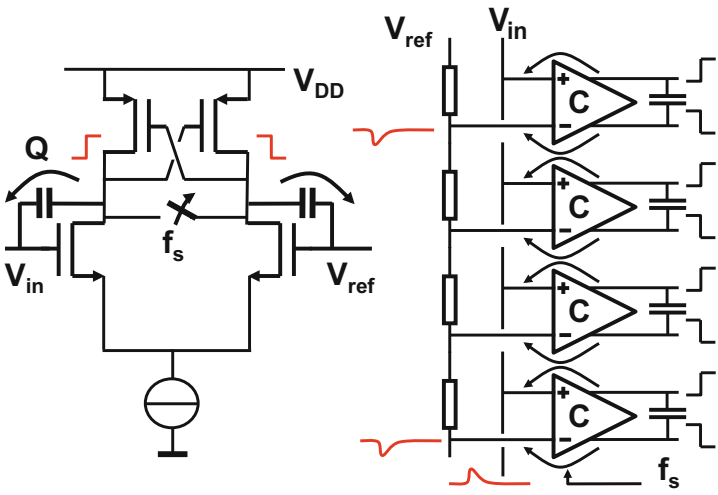


Fig. 8.10 Kick-back in a comparator circuit causes currents in the input and reference structure

8.1.6 Kick-Back

The comparator is a non-linear element: the only relation between the input signal and the output is the sign. The decision process in the comparator is often implemented with transistors that pass through all operation modes of the devices. The charges controlling the devices will show considerable variation. These charges have to be provided and will be dumped preferably in the power supply lines, however also an exchange to the input and reference signals will occur.

Figure 8.10 (left) shows a differential input pair loaded with a latch. The input voltage difference will lead in the succeeding circuits to a decision marked by sharp voltage transitions. The drain-gate capacitors of the MOS devices forming the input pair will pass these fast edges to the input and reference nodes thereby generating "kick-back." In some comparator designs these transitions can be as large as the power supply voltage. In this topology the drain of one input device will be pulled to V_{DD} and remain in inversion, while the current of the other device will go to zero.

Its inversion charge will flow either into the terminal, either the reference or input. This will cause even a larger kick-back effect.

Figure 8.10 (right) shows a flash circuit. In this architecture the kick-back effect is strong as the contributions of many comparators add up. During the simulation of the comparator ideal voltage sources instantly drain the charges from saturation effects and therefore will result in an optimistic performance prediction. Realistic impedances at the input terminals of the comparator allow a better simulation prediction of the performance in the presence of kick-back.

Reducing the kick-back signal requires to decrease the coupling capacitances by reducing the dimensions of the input related circuit elements with a penalty of higher random offset. Inside the comparator the swing of the signals must be sufficient to drive the next stages, but can be kept as low as a few tenths of a Volt by swing limiting measures. The additional advantage is that the transistors in the comparator remain in inversion, thereby avoiding time loss due to recharging. Some authors [287] apply dummy switches in analogy to Fig. 3.6, a solution that comes with the issues mentioned in that paragraph. A rigorous method applies an additional pre-amplifier. Unfortunately an additional stage often compromises the (speed, power) performance and will create additional delay.

8.1.7 Hysteresis

A comparator is designed to discriminate between positive and negative levels at its input terminal, irrespective of how small these levels are and what the previous decision was. In many comparator designs this ideal situation is not achieved. Either intentionally or as an unwanted consequence of the topology, the comparator remembers its previous state. Figure 8.11 shows an elementary comparator consisting of two inverters. The output of the second inverter is partly fed back in positive phase to the input. The comparator input has now a preference to keep its present

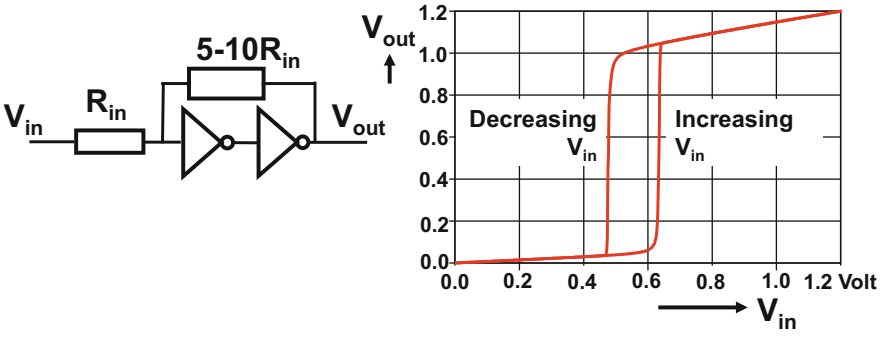


Fig. 8.11 Hysteresis created by positive feedback in a two-stage buffer

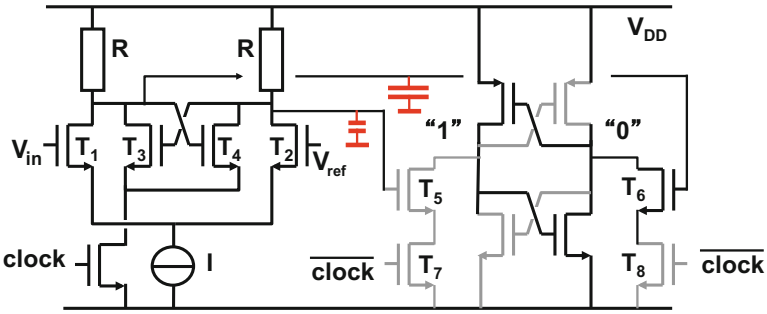


Fig. 8.12 Hysteresis created by the signal-dependent loading of the input latch by the second latch. The *grey-colored* transistors are not conductive and form only a small load capacitor

state and a small input value around the trip level of the inverter will not cause the output to change. Depending on the resistor ratio, the input will see a trip level that depends on the present state of the output. Going from negative input to positive or vice-versa results in different switching characteristics, see Fig. 8.11 (right). This effect is in analogy with magnetic materials called "hysteresis". In this example the hysteresis creates an additional threshold for change. The example circuit in Fig. 8.11 belongs to the class of "Schmitt-trigger" circuits. These comparators use the hysteresis threshold to avoid unwanted transitions in case of noise signals. In other words, a Schmitt-trigger circuit decides only if a clear input signal is present and the hysteresis avoids that noise generates false outputs.

A practical example of unwanted hysteresis is a pre-latch stage that is connected to a second stage consisting of a full digital latch activated by the inverse clock. In Fig. 8.12 the signal is fed from the amplifier to the full latch via two transistors T_5, T_6 that are in series with the inverse-clocked transistors T_7, T_8. The left-hand transistors T_5, T_7 will be out of inversion when that side of the latch stores a "1" signal. The connection to the gate of T_5 sees only some parasitic loading. The right-hand transistor T_6 sees a "0" signal and its channel is in full inversion. The inversion layer is connected to ground level creating a full gate-capacitor load. Thereby the two coupling transistors T_5, T_6 create an unequal capacitive loading of the first stage. This loading favors a similar signal in the first latch when the latch sees only a small input signal. The consequence is a comparator with hysteresis.

In other configurations the opposite effect may be the case: the hysteresis favors change and the comparator toggles at every clock cycle for an input signal close to the trip level.

The previous state can be removed by resetting the latch in the comparator. Some designs simply disconnect the latch from the supply. This method assumes that there is sufficient time to discharge the latch nodes. A more active approach consists of adding a specific clear or reset mechanism to the latch. Figure 8.13 shows two configurations. The left-hand circuit connects the latch nodes to the power supply and actively removes all state information. The low-ohmic switches change the

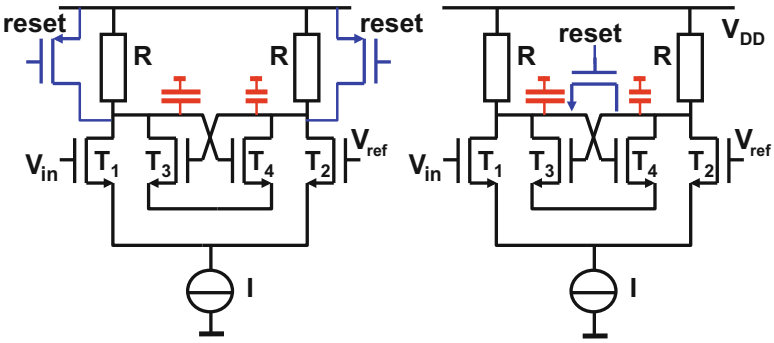


Fig. 8.13 Two methods to clear the state in comparator circuit

DC-levels in the amplification branch. The DC-level must be restored and during that process any inequality in capacitive loading will show up as inaccuracy.

The construction on the right-hand side circumvents the restoration problem by pinching the two latch nodes together. This will keep the average DC-level of the nodes in tact and sets the latch to its trip point. A proper choice of the impedance of the switch allows to use it as a differential load for the current of the input pair. The reset transistor must be bootstrapped in case of low power supply.

8.1.8 Comparator Schematics

Many comparator designs have been published in literature and there is obviously not a standard choice for a comparator circuit topology. This is, on one hand, caused by the evolution of the technology. The drive capabilities of the transistors have improved, but the power supply voltage and the input signals have reduced. On the analog-to-digital converter architecture side also a lot of developments have taken place over the last 25 years. The rise in popularity of the pipeline converter has allowed to spend more area and power on a comparator than a flash converter can tolerate. The following examples of comparator circuits are a subset of the published work and are meant to educate and perhaps inspire a new design.

The comparator published in [180] was designed in a 7 μm NMOS metal-gate technology. Figure 8.14 shows a CMOS variant of that design. A similar topology is found in [163, Fig. 13]. The input stage serves to amplify the differential signal and allows some common mode rejection. The second stage is fed with the amplified signal to generate a digital decision. The intermediate amplification nodes separate the pre-amplifier from the latch. The problem with an intermediate node is that it creates an additional pole and slows down the design. In the design of Fig. 8.14 the poles are formed by the PMOS mirrors. These mirrors cannot easily be operated at a large gate-source voltage as the input range is reduced. This mirror also contributes to the input random offset. The strict separation of latch and input,

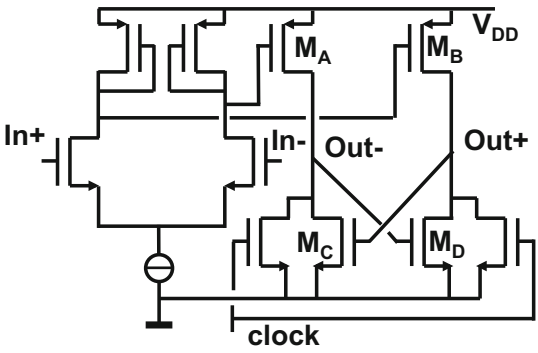


Fig. 8.14 A comparator in CMOS based on an NMOS enhancement/depletion design [180]

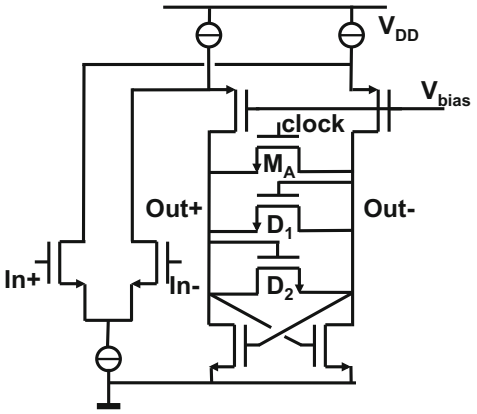


Fig. 8.15 A comparator designed for a folding analog-to-digital converter [181]

however, allows to minimize any kick-back effect. This design is suited for input and reference sources that cannot tolerate kick-back. Obviously both branches continuously consume DC power.

The original design (1981) resulted in a random offset at the input of 6.1 mV, 5 bit resolution, 4 MHz signal bandwidth at 20 Ms/s.

The comparator in Fig. 8.15 is designed for a folding analog-to-digital converter [181]. It consists of an input stage that feeds its differential current in a cascode stage. This construction avoids an intermediate mirror pole that would slow down the circuit. Still a reasonable shielding for kick-back is obtained as long as the cascode remains functional. In order to guarantee that, the latch is equipped with two diode-connected transistors to have a current path available even if the latch is activated. Saturation of the cascode and input stages is thereby prevented. The current consumption is significant due to the current sources. In a 0.8 µm CMOS process the 1-sigma input referred random offset is 2.5 mV. The circuit runs up to 70 Ms/s sample rate in this process. Its speed will certainly benefit from transfer to advanced processes.

Figure 8.16 uses three phases to reach a decision [182]. When inverse clock is high the comparator latch is in reset mode. The signal is amplified over the resistance of the reset transistor. After the inverse clock goes low, during a short

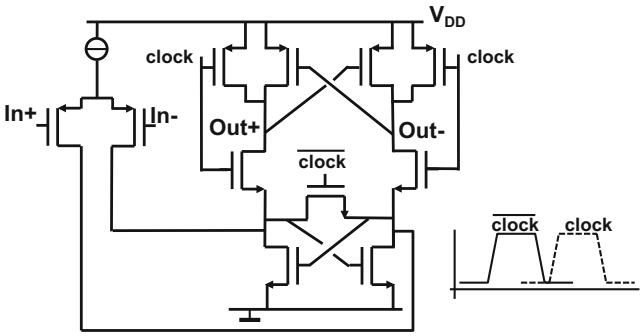
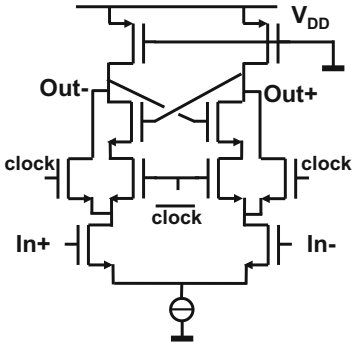


Fig. 8.16 A comparator aiming at high-speed operation [182]

Fig. 8.17 A comparator designed based on a former bipolar design [183]



period the lower latch can amplify the signal as if it were an integrating pre-amplifier. This amplification allows to suppress the mismatch of the upper latch section. After the clock signal goes high, this pre-amplified signal is boosted to digital levels. Some sensitivity to the non-overlapping clock phases is present, as well as a considerable kick-back. The current source consumes continuously power.

The comparator in Fig. 8.17 is based on former bipolar designs [183], where dominantly npn-transistors and resistors are used. Here the NMOS transistors replace the npn devices and PMOS transistors in their linear region act as resistors. The comparator routes a constant current through either the outer amplification side (clock is high) or through the inner latch transistors. The speed is limited by the latch pair and its load. A lot of device imperfections add up in the current path, therefore a higher input referred mismatch is expected. Also the kick-back charge can be significant. This effect will depend on how well the drains of the input pair are kept stable during clock transition. The accuracy of the crossing of the clock and the inverse clock is crucial. The design was used in a 0.5 μm CMOS process with a clock speed of 80 Ms/s.

A comparator for a 4-bit 12 Gs/s analog-to-digital converter uses a two-stage comparator, Fig. 8.18. The middle PMOS transistor M_1 in the pre-amplifier serves as a load for the input pair. The sources of the input pair are switched to the positive

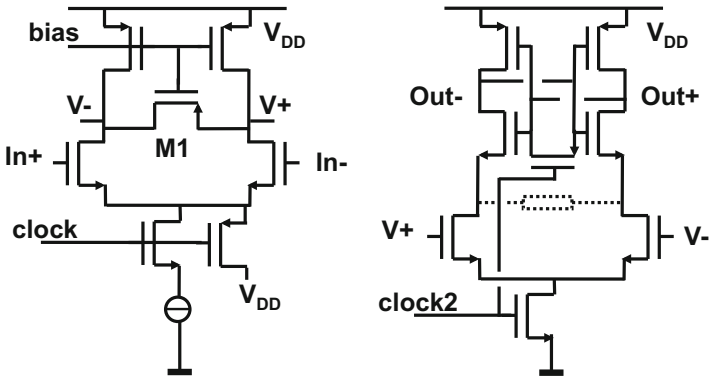


Fig. 8.18 A comparator for high speed operation [184]. The latch (*right*) is based on a design for the "StrongArm" processor [185, 186]

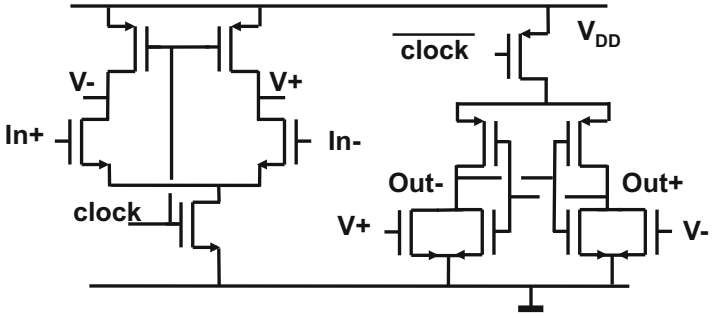


Fig. 8.19 A double-tail comparator combines accuracy with speed [188]

power supply at the end of the amplification, which produces kick-back charge via the coupling of the gate. The latch(right) is also used in memory and interface circuits [185] and in the "StrongArm" design [186] and is not intended to resolve small voltage differences as it serves as an edge-triggered latch. If the clock is low, the cross-coupled pair is reset and no current flows. A high-ohmic resistor (dotted) prevents leakage currents to charge nodes asymmetrically. At rising clock the input signal is amplified and regenerated in the latch to a full digital signal. After full settling the latch is consuming no current. The 1-sigma input referred random offset in [184] was in 0.25 µm CMOS 60 mV and the bandwidth exceeded 2.5 GHz at 12 Gs/s. Another example of an implementation of this comparator is in a 5-bit flash converter [187] where a good energy efficiency is achieved after calibration of each comparator.

The comparator in Fig. 8.19 [188] is a further development of Fig. 8.18. The amplification is now carried out in an integrating pre-stage, see Fig. 8.5, thereby reducing the amount of stacked transistors and the kick-back noise. The design does not require any DC current, which makes it very suited for achieving low-power

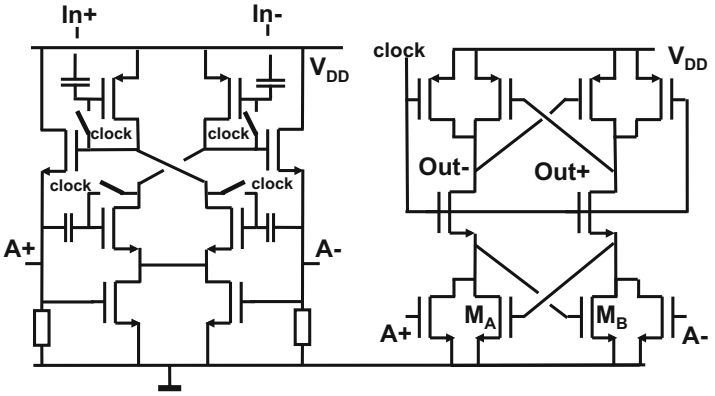


Fig. 8.20 A comparator with capacitive common mode based on the work in [189]

operation. The design in 90-nm CMOS achieves 2 Gs/s, $\sigma_{in,offset} = 8$ mV, $\sigma_{in,noise} = 1.5$ mV with 90–110 fJ/decision power consumption. The internal time constant of the comparator can be estimated at 19 ps.

The above comparators are all based on a differential input pair that some common mode voltage range. The comparator in Fig. 8.20 also consists of a preamplifier(left) and a latch [189]. The preamplifier does not use a differential pair with tail current, but the input voltages are capacitively coupled into the comparator. This configuration allows a very wide range of input and reference voltages. In the cross-coupled loop formed by the source followers and the capacitors, another capacitive voltage shifting is used. The necessary pre-charging of the capacitors will create some kT/C noise. The right half of the circuit consumes only power during transitions, the power in the left-half is significant. The residual offset is $\sigma_{in,offset} = 2$ mV and in a 2 µm process 40 Ms/s was reported.

A more advanced comparator with the same idea is shown in Fig. 8.21. The design is differential, thereby eliminating the PSRR problems. The random input offset of the comparator is estimated at $\sigma_{in,offset} = 36$ mV [39]. Still some time is needed for the bias setting in this 0.18 µm CMOS design. Interleaving of two input stages reduces in this converter this speed penalty.

Example 8.2. A differential NMOST input pair (W/L = 50/2) is loaded with a PMOST current mirror (W/L = 36 µm/1 µm) in a process with $A_{VT,N} = A_{VT,P} = 6$ mVµm, and the ratio between the NMOS and PMOS current factor for W/L = 1 is $\beta_{\square,n}/\beta_{\square,p} = 3$. Calculate the input referred mismatch.

Solution. In Fig. 8.22 the schematic of this circuit is drawn. With the help of Eq. 5.16 the standard deviations for the NMOS and PMOS threshold voltages are found: $\sigma_{VTN} = 6/\sqrt{50 \times 2} = 0.6$ mV, and $\sigma_{VTP} = 1$ mV. The PMOS transistor mismatch must be referred back to the input. This threshold mismatch translates into a PMOS current mismatch via: $g_{m,P}\sigma_{VTP}$. The effect at the input terminals is the input referred mismatch voltage. This voltage must generate in the NMOS

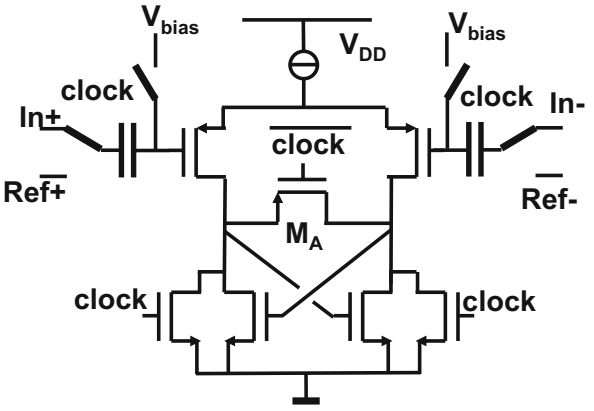


Fig. 8.21 A comparator design for an interleaved pipeline converter [39]

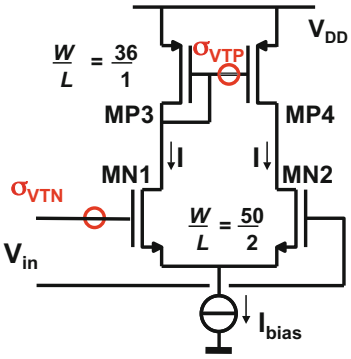


Fig. 8.22 The circuit schematic

transistors a current that compensates the mismatch current generated by the PMOS devices: $g_{m,N}\sigma_{VinP} = g_{m,P}\sigma_{VTP}$. So the total input referred mismatch is composed of the NMOS and PMOS contributions:

$$\sigma_{VinTOT} = \sqrt{\frac{g_{m,P}^2}{g_{m,N}^2}\sigma_{VTP}^2 + \sigma_{VTN}^2} = 0.92\,\text{mV}$$

8.1.9 Calibrated Comparators

Input referred random offset is one of the key problems to battle in comparator design. Early on attempts have been made to "auto-zero" [190] or cancel [191] this offset. The basic schematic of Fig. 8.23 is similar to the track-and-hold circuit in Fig. 3.22. During the auto-zero mode the switches are in the dashed positions and the inverter is essentially in unity gain mode. The capacitor is charged on its left hand

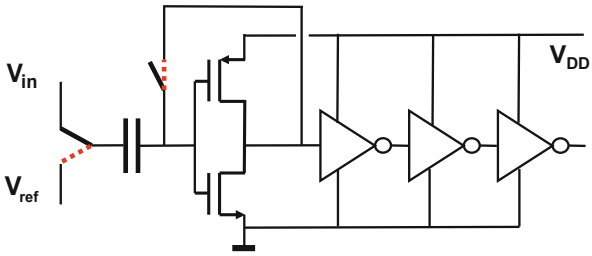


Fig. 8.23 An auto-zero comparator as used in [190]

side to V_{ref} and on the right-hand side to the input offset. When the switches toggle the change on the input of the comparator equals $V_{in} - V_{ref}$. This voltage is amplified with the small-signal gain of the inverter and its successors. Some residual offset is due to limited gain and charge variation in the switches. This scheme was used in a 1.4 µm CMOS process [190] and allowed an analog-to-digital converter running at 40 Ms/s with 10 MHz bandwidth. It is clear that the charging and discharging of the input capacitor puts extreme demands on the input and the reference impedances. The disadvantage of this method is that the high capacitive input load requires a low-ohmic ladder (300 Ω) and poses high demands on the circuitry for driving the analog-to- digital converter. The single-sided input capacitor has a large parasitic capacitance to a bouncing substrate in a mixed signal chip. The implementation as depicted in Fig. 8.23 has a poor power supply rejection [192], which affects the performance if power supply bouncing occurs between various conversion cycles. A large part of the power supply variation is translated in a signal contribution on the input. In the case of an inverter the PSRR is around 0.5. So the power supply has to be kept clean to the level of a few V_{LSB}. These comparator design points have to be solved for embedded operation.

Also the timing of the offset-cancellation is a point of consideration. Offset compensation per clock cycle costs 30–50 % of the available sample period and reduces the maximum speed. Offset compensation at a fraction of the sample rate is possible, but can introduce spurious components at those lower frequencies.

The auto-zero cancellation, as shown in Fig. 8.23, acts as a transfer function for all (unwanted) signals $e(z)$ that originate at the input of the inverter. The error component in the output signal is found in a similar manner to the analysis in Sect. 3.4. The error component in the output signal is

$$V_{out,error} = e(z)(1 - z^{-0.5}) \tag{8.10}$$

The exponent −0.5 assumes a 50 % duty cycle of the switching signal. This transfer is characterized by a term $2\sin(\pi f/2f_s)$. Low frequency components are suppressed, but higher frequencies can even experience some amplification. Moreover the switching sequence acts as a sampling mechanism for unwanted high-frequency signals thereby shifting these high frequencies into the signal band. The auto-zero capacitor acts as a sampling capacitor for which the kT/C noise analysis applies.

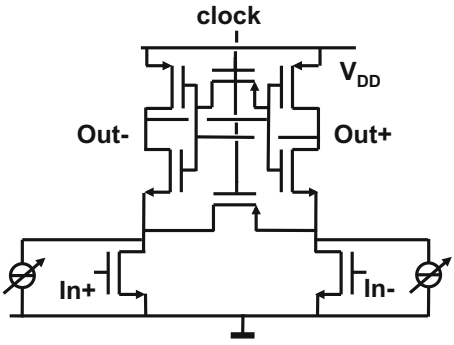


Fig. 8.24 The current sources are controlled to auto-zero this comparator [193]

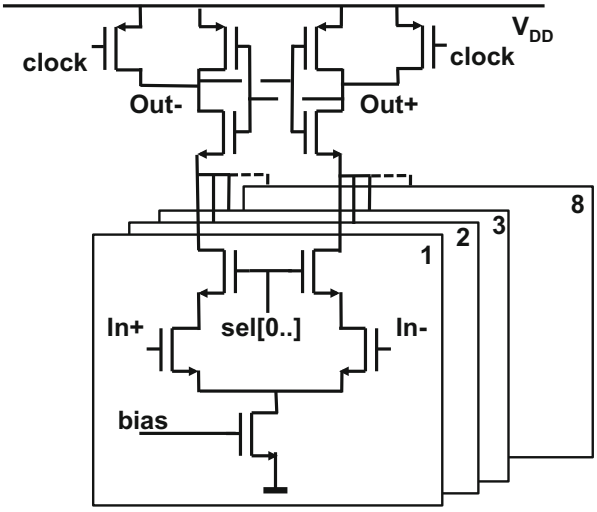


Fig. 8.25 In order to configure the input differential pair for low offset four out of eight available transistor pairs are selected [194]. A similar topology is used in [287]

Running this offset compensation scheme at low frequencies results in stacking of noise and down-sampling of spurious signals (e.g., from the substrate). See also the current-calibration technique in Sect. 7.6.

Another class of auto-zero mechanisms uses adjustable voltages or currents to reduce the input referred offset. Figure 8.24 shows the basic idea implemented with two controlled current sources [193]. The necessary corrections are carried out by means of an algorithm, either at start-up or during operation. If the chip contains non-volatile memory also programming after production is possible.

An interesting alternative to the above methods uses multiple parallel input differential pairs [194] of minimum dimension. In the example of Fig. 8.25 four out of eight available input pairs are chosen to match the required input transconductance. Statistics tells that there are 70 possible combinations choosing four elements from a total of 8. The combination with the lowest joint offset is selected. The required logic

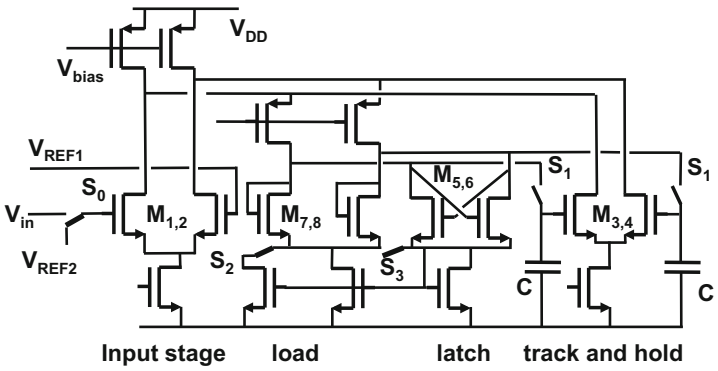


Fig. 8.26 A multi-purpose comparator schematic: it performs track and hold, comparison and reference subtraction. Switches are shown in the position for the amplification phase. S_3 activates the latch. Design: J.v. Rens

is tolerable in advanced CMOS processes. This comparator was used for a time-interleaved 6-bit 20 Gs/s analog-to-digital converter in a 32-nm SOI technology.

8.1.10 Track-and-Hold Plus Comparator

Random-offset reduction is necessary to achieve better than 8-bit DNL performance. In most offset reduction schemes the offset+signal and the offset are determined at different points in time, so at least one must be stored in a capacitor. Figure 8.26 shows the comparator that was used in several analog-to-digital converter designs (see Sect. 8.6.7).

With the exception of input and ladder terminals, the design is fully differential and the PMOS current sources allow a good PSRR. The design consists of an input stage $M_{1,2}$, from which the signal is fed into a sample-and-hold stage $M_{3,4}, S_1, C,$. Comparison involves three cycles: sampling, amplification, and latching. During the sampling phase switch S_1 is conducting and $V_{in} - V_{ref} + V_{off}$ is stored on both capacitors. These capacitors are grounded on one side and do not suffer from parasitic coupling to substrate. V_{off} represents the sum of all the offsets in the comparator. During the sampling phase the negative conductance of latch stage $M_{5,6}$ is balanced by the positive conductance of the load stage $M_{7,8}$. Their combination acts as an almost infinite impedance, which is necessary for good signal+offset storage. The latch stage has twice the W/L of the load stage, but its current is only half due to the current mirror ratio. For large differential signals the effect of the factor 2 in W/L of $M_{5,6}$ and $M_{7,8}$ becomes important: the conductance of $M_{5,6}$ reduces at a much higher rate that of $M_{7,8}$, so the effective impedance of $M_{5,6,7,8}$ collapses. The large-signal response is consequently improved by the reduced time

constant. The feedback of $M_{3,4}$ via the switches S_1 allows a 150 MHz bandwidth in a 1 µm technology, resulting in a high-quality T&H action.

After the sampling phase, the switch S_0 at the input connects to the reference voltage, effectively disconnecting the common input terminal from the comparator. As switches S_1 are disconnected, the sample-and-hold stage will generate a current proportional to $V_{in} - V_{ref} + V_{off}$, while the input stage and the rest of the comparator generate only the part proportional to V_{off}. The (differential) excess current is almost free of offsets and will be forced into the load-latch stage $M_{5,6,7,8}$. Switch S_2 is made conductive, which increases the conductance of $M_{7,8}$ and decreases the conductance of $M_{5,6}$; the gain of $M_{3,4}$ on $M_{5,6,7,8}$ is now about 8 for small differential signals.

Finally, S_3 is made conductive; the current now flows in a 2:1 ratio into $M_{5,6,7,8}$, thereby activating the latch operation. The latch decision is passed on to the decoding stage and a new sample can be acquired. Remaining random offsets (approx. 0.4 mV) are caused by limited gain during the sampling phase, charge dump of S_1 and the difference in matching contributions of $M_{5,6,7,8}$ in the sampling and amplification phases.

8.2 Flash Converters

Flash[6] converters are used for two main purposes. As a stand-alone device, this converter can achieve the highest conversion speeds for low (6-bit) resolutions [195]. Another important field is the application in lots of other analog-to-digital architectures, such as subrange converters.

A flash converter is comprised of a resistor ladder structure whose nodes are connected to a set of $2^N - 1$ comparators, see Fig. 8.27. A decoder combines the comparator decisions to a digital output word.

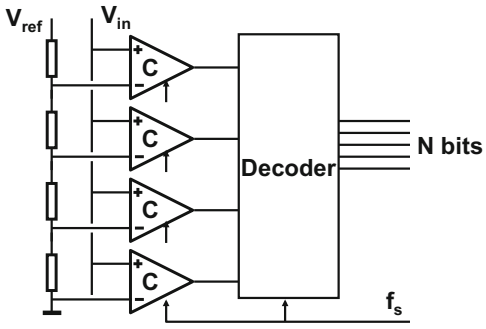


Fig. 8.27 A "flash" analog-digital converter

[6]The origin of the addition "full" in full-flash can refer to the conversion of the full range. "Partial-flash" converters can refer to a subranging architecture.

The input signal is compared to all reference levels simultaneously. At the active sampling clock edge, a part of the comparator circuits with an input signal lower than the local ladder voltages will generate a logical "zero," while the other comparators will show a logical "one." The digital code on the outputs of the comparators is called a "thermometer code." A digital decoder circuit converts the thermometer code in an N-bits output format (mostly in straight binary format).

The input signal for a flash converter is only needed at the moment the latches are activated. The sample-and-hold function is inherently present in the digital latches of the comparators. For most applications an external sample-and-hold circuit is not needed. Flash converters are the fastest analog-to-digital converters, however their complexity and their power consumption grow with the number of comparator levels 2^N. Also the area and the input capacitance increase exponentially with N.

The capacitance at the input of a flash converter is largely determined by the input capacitance of the comparator. If that input capacitance is formed by, e.g., a differential pair, the effective input capacitance of the comparator will depend on the input signal. An input signal lower than the local reference voltage of the comparator will result in a current starved input branch of the differential pair. The signal will experience a low input capacitance as the MOS input transistor is out of inversion. A high input signal creates an inversion charge, and a considerable input gate-source capacitance creating a high input capacitance. The group of comparators below the decision level create a high input capacitance and the group above the decision level will show a low capacitance, Fig. 8.28. The design of the comparator will affect the magnitude of the input capacitance variation, yet most comparator designs result in an input impedance of a flash converter that is (somewhat) signal dependent. Second-order distortion will arise if the converter is fed from a source with source impedance. The input capacitance for single-sided operation is found as:

$$C(V_C(t)) = C_0 + \Delta C \frac{V_C(t)}{V_{ref}} \tag{8.11}$$

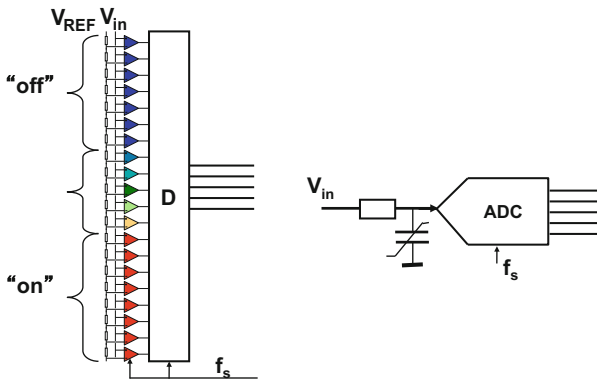


Fig. 8.28 A flash converter presents a non-linear input impedance

where ΔC represents the total capacitive change for an input-voltage swing from 0 to V_{ref}. V_C is the voltage over the non-linear capacitor that holds a charge

$$Q_C(t) = \int C(V_C(t))dV_C(t) = C_0 V_C(t) + \Delta C \frac{V_C^2(t)}{2V_{ref}}$$

Now the capacitive current is found:

$$I_C(t) = \frac{dQ_C(t)}{dt} = C_0 \frac{dV_C(t)}{dt} + \frac{V_C(t)\Delta C}{V_{ref}} \frac{dV_C(t)}{dt} \tag{8.12}$$

A correct analysis requires to equate this current to the current through the resistor, resulting in a non-linear differential equation. The substitution $V_C(t) = 0.5V_{ref} + 0.5V_{ref}\sin(\omega t)$ assumes that the non-linear term is relatively small and will not substantially change $V_C(t)$. Evaluation of the last part of the equation leads to a second harmonic current, which is multiplied with the input resistor R and compared to the first harmonic term in $V_C(t)$:

$$\text{HD2} = \frac{\omega R \Delta C}{4} \tag{8.13}$$

This estimate for HD2 is proportional to the signal frequency, the input impedance and the amount of capacitive variation. For $f = 1\,\text{GHz}$, $R = 50\,\Omega$ and a capacitive variation of 100 fF a distortion level of $-42\,\text{dB}$ results. Changing the comparator into a full-differential design, with two input terminals and two reference voltage terminals will solve this problem at the expense of more current and area.

Flash converters are predominantly used in high-speed applications. These applications require tight control on the timing of the converter. Jitter control of the sampling pulse is crucial to high performance, see Sect. 2.6.1. However, also in the actual design of the converter attention has to be paid to avoid unbalanced timing pulses in the circuit. The impedance of the wiring in combination with a string of load elements (e.g., inputs to the comparators) can easily lead to delay differences between individual comparators. In the example of Fig. 8.29 (left) the clock and the signal come from opposing sides in the structure. For signal levels close to the bottom of the structure the signal will be delayed with respect to the sample clock, while at levels close to the top the signal will be advanced with respect to the sample clock. In a structure where the signal conversion is linearly related to the position of the comparators, the relative delay (ΔT) of the signal versus the clock is proportional to the signal.

$$\begin{aligned} V_{in}(t + \Delta T(t)) &= V_a \sin(\omega(t + \Delta T \sin(\omega t))) \approx V_a \sin(\omega t) + V_a \omega \Delta T \sin(\omega t)\cos(\omega t) \\ &= V_a \sin(\omega t) + 0.5 V_a \omega \Delta T (1 - \cos(2\omega t)) \end{aligned} \tag{8.14}$$

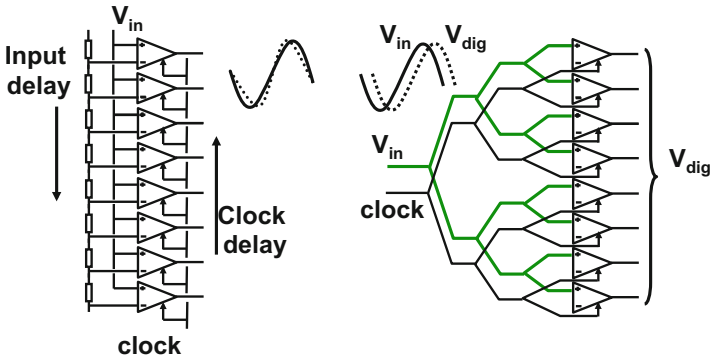


Fig. 8.29 A simple distribution strategy for input signal and clock can easily lead to performance loss at high frequencies (*left*). A tree-like lay-out is necessary to avoid delay differences

This timing error translates in a second order distortion component with a relative magnitude of $0.5\omega\Delta T$. Even with modest signal frequencies of, e.g., 10 MHz and a delay of 100 ps this results in a −50 dB distortion component. This example shows that the relative timing of the signal and sample must be accurate. Therefore a tree-like structure as in Fig. 8.29 (right) is applied in high performance converters.

Example 8.3. A group of 64 comparators is on regular intervals connected to the input signal line of total length 640 μm and width 1 μm. The square resistance is 0.1 Ω. Every comparator has an input capacitance of 0.1 pF. The clock arrives synchronously at all comparators. Estimate the expected distortion for a 100 MHz signal fed into one side of the line.

Solution. The total resistance of the wire is $(640/1) \times 0.1\,\Omega = 64\,\Omega$. The total capacitance is 6.4 pF. In Sect. 7.2.1 an analysis was made for a double terminated structure with distributed resistance and capacitance. The structure in the present example can be considered is only connected on one side. The problem can be mapped on Sect. 7.2.1 by considering a twice as long line. Exactly in the middle there is for symmetry no current flowing and the time constant will not change a half of that problem. So the time constant is easily found by considering that a double structure behaves with $\tau = R_{tot}C_{tot}/\pi^2$ leading to $\tau = 0.167$ ns. The relative magnitude of the second harmonic component is estimated with Eq. 8.14 as 5.2 %.

8.2.1 Ladder Implementation

In Sect. 7.2.2 the dynamic behavior of a resistor string was analyzed. This theory is applicable to the design of the ladder structure in the flash converter. The time-constant for settling was found to be

$$\tau = rcL^2/\pi^2 \tag{8.15}$$

with r and c the resistivity and capacitance per unit length. In a flash converter with N bit resolution this equation is rewritten as

$$\tau = R_{tap} C_{tap} 2^{2N} / \pi^2 \tag{8.16}$$

R_{tap} is the resistance between two tap positions on the resistor ladder and C_{tap} represents the total capacitance on a tap and is composed of the input capacitance of the comparator, the bottom-plate stray capacitance of the resistor and all wiring parasitics. With $C_{tap} = 0.1\,\text{pF}$, $R_{tap} = 1\,\Omega$ and $N = 7$ a time constant $\tau = 0.16\,\text{ns}$ will result, which would allow a sampling speed of around 1 Gs/s. The ladder impedance is 128 Ω. A 1 V reference voltage over the ladder already requires 8 mA.

In a similar manner as the time constant, the deviations in ladder voltages due to DC-current that, e.g., bipolar transistors or some auto-zeroing schemes can draw from the ladder, can be estimated. See Example 7.2 on page 226:

$$\Delta V_{middle} = R_{tap} I_{tap} 2^{2N} / 8 \tag{8.17}$$

With the same parameters and $I_{tap} = 10^{-6}\,\text{A}$, the voltage deviation in the middle of the ladder equals 1.6 mV. This loading effect of the ladder by the comparators is more pronounced in bipolar design. Darlington stages in the comparators are used to reduce the loading of the ladder by base currents and at the same time reduce the kick-back effects [196].

This example shows that even modest specifications in a standard flash converter require a low-ohmic ladder. Often the ladder is constructed from the metal layers or special materials on top of the device. These materials can easily exhibit gradients and although no more than 6–7 bit of resolution is required, some precautions have to be taken to mitigate the gradients. An anti-parallel connection of two ladders reduces the gradient as in Fig. 5.10.

8.2.2 Comparator Random Mismatch

The key requirements for the choice of a comparator are a low capacitive load, no DC input current, low random offset, low kick-back, high switching speed and bandwidth, and low power. Most of these requirements can be met by using small size transistors, with the exception of low random offset.

One of the major differences in the design of CMOS and bipolar analog-to-digital converters is the lack of accuracy in CMOS comparators. A bipolar differential pair has a random offset on the base-emitter voltage V_{be} in the order of $\sigma_{\Delta Vbe} = 0.3\,\text{mV}$. A pair of NMOS transistors with small gate lengths show random offsets in the order of $\sigma_{\Delta VT} = 2\text{–}6\,\text{mV}$, see Sect. 5.3. Because of the low CMOS gain, the offsets in the remaining transistors of the comparator contribute too, which may lead to $\sigma_{in} = 5\text{–}20\,\text{mV}$ as a total input-referred random offset. Obviously the LSB-size must exceed $3 \times \sigma_{in}$, limiting the overall resolution to 4–5 bit.

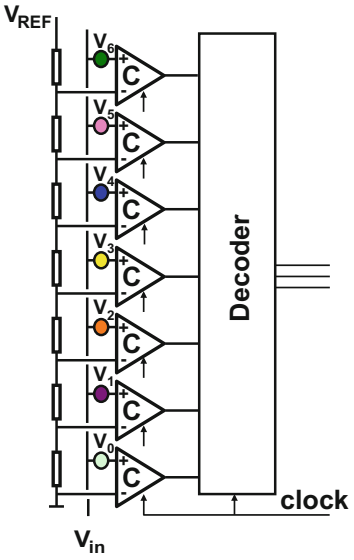


Fig. 8.30 A basic 3-bit flash analog-to-digital converter is limited in its performance by the input referred mismatch of the comparators

The random comparator offset in an N-bit flash analog-to-digital converter (with $2^N - 1$ comparators, see Fig. 8.30) affects the DNL or non-monotonicity (DNL $= -1$ LSB). The DNL for a converter with 2^N conversion levels is specified as

$$\mathrm{DNL}_j = \frac{V_{j+1} - V_j}{V_{LSB}} - 1 \quad \mathrm{DNL} = \mathrm{Max}(|\mathrm{DNL}_j|)\ \forall j = 0, \ldots, 2^N - 2 \tag{8.18}$$

V_j is the value of the input signal which causes comparator j to flip and consequently contains the comparator random input referred offset. V_{LSB} is the physical value corresponding to an LSB.

The measured DNL of a converter during final test is often used as an elimination criterium and thereby determines the yield. The actual value of the DNL, being the maximum of the absolute value of $2^N - 1$ differences of stochastic variables, is a random function itself. Figure 8.31 shows the distribution of the actual DNL of 1000 8-bit flash converters, generated by means of Monte-Carlo simulation. This graph represents 255×1000 comparators with an input referred mismatch characterized by $\mu_{in} = 0, \sigma_{in} = 0.15 V_{LSB}$. The shape of the distribution is characteristic for production measurements of various types of converters. The distribution is not Gaussian as the DNL is a non-linear function. The comparator random offset was in this example chosen to be $\sigma_{in} = 0.15 V_{LSB}$, corresponding to $\sigma_{in} = 0.586 \times 10^{-3} V_{ref}$. There are almost no trials with an actual DNL < 0.5 LSB, although all trials result in monotonicity. For half of the trials the actual value of the DNL is lower than 0.64 LSB, which is also found in the first-order calculation of Example 8.4.

Example 8.4. Estimate the DNL of one 8-bit flash analog to digital converter with $\sigma_{in} = 0.15 V_{LSB}$.

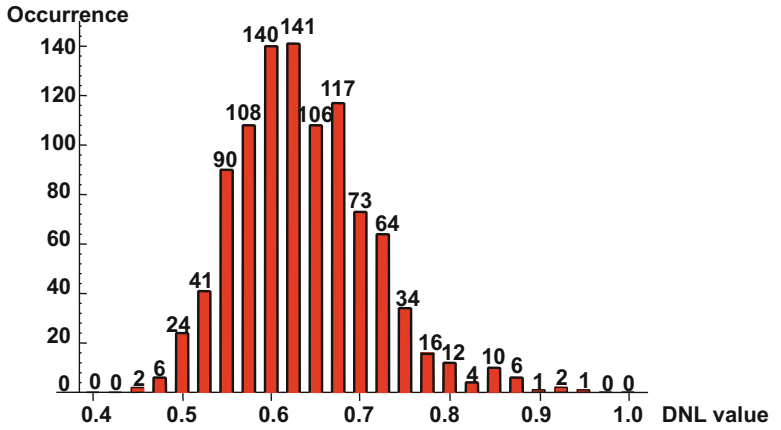


Fig. 8.31 Typical histogram of DNL values for an 8-bit flash architecture, with $\sigma_{in} = 0.15V_{LSB}$. The mean DNL in this simulation is 0.64 LSB

Solution. The DNL_j set by the j-th and $j+1$-th comparator in a flash converter is given by $\text{DNL}_j = (V_{j+1} - V_j)/V_{LSB} - 1$ and is normalized to (expressed as a fraction of) V_{LSB}. Its standard deviation equals $\sqrt{\sigma^2_{in,j+1} + \sigma^2_{in,j}}$. As all input referred random offset are modeled with the same normal distribution $N(0, \sigma^2_{in})$, the standard deviation of DNL_j is $\sigma_{in}\sqrt{2} = 0.15\sqrt{2}V_{LSB}$. All DNL_j of the entire converter form therefore also a normal distribution $N(0, 2\sigma^2_{in})$.

The overall DNL or the DNL of the complete analog-to-digital converter is the absolute value of the most extreme value of this distribution. For that extreme value, it must hold that there is a good probability that all DNL_j are within that extreme value. If an excess factor in Table 8.2 of $z = 3$ is used, then the probability that one DNL_j exceeds this value is $2 \times 0.00135 = 0.0027$. On the level of the converter all 254 of these DNL_j's must fulfill this requirement, so there the yield is: $(1 - 0.0027)^{254}$=0.5, or 50 % yield. That excess factor is an estimate of the overall DNL: $z \times \sigma_{in}\sqrt{2} = 3.0 \times 0.15\sqrt{2}V_{LSB} = 0.64V_{LSB}$. So the expected DNL is 0.64 LSB. Note that infact the DNL of a large population of converters is a distribution function as in Fig. 8.31.

More formally:

$$\begin{aligned} \textit{Yield} = (1-p)^{2^N-2} < 0.5 \qquad & \rightarrow p > 0.9973 \\ \textit{from Table 8.2} \quad p = 0.9973 \leftrightarrow \alpha = 3.0 \qquad & \rightarrow \frac{\text{DNL} \times \text{V}_{\text{LSB}}}{\sigma_{in}\sqrt{2}} = 3.0 \\ \text{with } \sigma_{in} = 0.15V_{LSB} \qquad & \rightarrow \text{DNL} = 0.64 \end{aligned} \tag{8.19}$$

A more formal analysis results in a yield prediction. If the input random offset of every comparator is given by a Gaussian distribution with zero-mean and a standard deviation σ, then the probability of all the comparators being within the

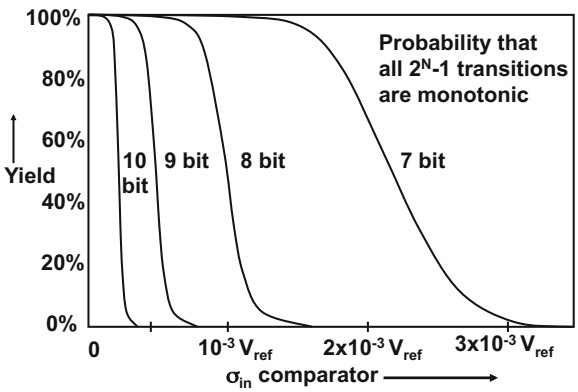


Fig. 8.32 Yield on monotonicity versus the standard deviation of the comparator random offset

monotonicity limit can be calculated. This probability is an estimation of the yield of the analog-to-digital converter. Ideally, $(V_{j+1}-V_j)-V_{LSB}=0$. Non-monotonicity occurs when comparator $j+1$ (fed with a rising input signal) switches before the adjacent comparator j with a lower reference voltage does. In mathematical formulation the probability that non-monotonicity occurs is: $p = P(V_{j+1} < V_j)$. Then $(1-p)$ is the probability of comparators j and $j+1$ switching in the correct order; this condition must hold for all 2^N-2 pairs of comparators, so:

$$\text{Yield} = (1-p)^{(2^N-2)} \quad \text{with}$$
$$p = P\left(\frac{V_{j+1}-V_j-V_{LSB}}{\sigma\sqrt{2}} < \frac{-V_{LSB}}{\sigma\sqrt{2}}\right) \tag{8.20}$$

The mean value of the argument in the probability function for p is 0, while the standard deviation of the argument corresponds to $\sqrt{2}\times$ the standard deviation of a single comparator σ in mV. The probability function is normalized to a standard normal curve N(0,1).

Figure 8.32 shows the curves that relate the yield to the standard deviation σ_{in} of the comparator random offset for an input voltage range of V_{ref}. A 10-bit converter requires a $\sigma_{in} < 0.25 \times 10^{-3} V_{ref}$. This is still achievable in bipolar technology [197]. For higher accuracies trimming, higher input voltages or other forms of offset correction are needed. The 8-bit converter from Fig. 8.31 with $\sigma_{in} = 0.15 V_{LSB} = 0.15 \times V_{ref}/256 = 0.586 \times 10^{-3} V_{ref}$ is indeed monotonic with close to 100 % yield. Indeed the largest DNL is in the 1000 trails is +0.94 or −0.94.

In Fig. 8.33 the requirements are more severe: now the probability of the converter achieving a DNL of less than 0.5 LSB has been calculated. The probability p of two adjacent comparators exceeding the DNL limits is

$$\text{Yield} = (1-p)^{(2^N-2)} \quad \text{with}$$
$$p = P\left(\left|\frac{V_j-V_{j-1}-V_{LSB}}{\sigma\sqrt{2}}\right| > \frac{\text{DNL}\times V_{LSB}}{\sigma\sqrt{2}}\right) \tag{8.21}$$

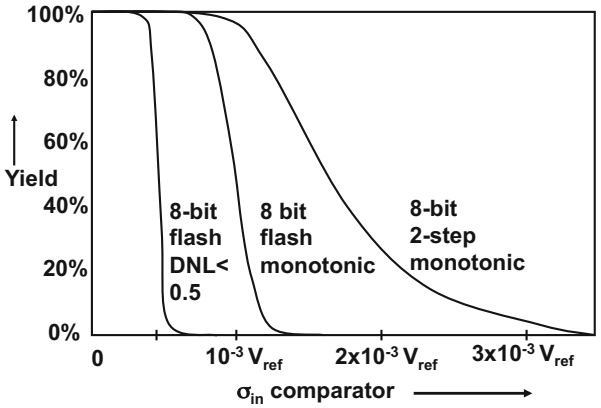


Fig. 8.33 Yield on DNL<0.5 LSB, on monotonicity for 8-bit flash and on monotonicity for an 8-bit 2-step subrange architecture

As expected, the σ_{in} required for a DNL value of 0.5 LSB has been more than halved with respect to the non-monotonicity requirement.

The demands for a two-stage subranging architecture are also shown for (a theoretical minimum of) 15 coarse and 15 fine comparators. Owing to the steep nature of the Gaussian distribution the advantage of having only 15 critical comparators results in marginally more tolerance on the input random offset voltage at a 95–99 % yield level. This example indicates that these yield considerations for flash architectures give a good first-order approximation for more complex architectures.

At first sight one bit more resolution requires that the comparator random offset should reduce by a factor of two. A 7-bit flash converter has to reach an input standard deviation of $1.8 \times 10^{-3} V_{ref}$ to achieve an acceptable yield. At the 8 bit level comparators with random offset better than $0.8 \times 10^{-3} V_{ref}$ are needed for the same yield. One bit more resolution translates in a factor two reduction of the random offset standard deviation. Moreover, the same yield must be reached with twice the number of comparators resulting in a random offset reduction factor of 1.8/0.8, slightly higher than 2. Table 8.1 shows the effect of increasing the resolution with one bit in a flash analog-to-digital converter. In this table it is assumed that the comparator's design comprises three transistor pairs that contribute to the random offset: the input pair, a current source pair, and a latch pair.

The first three lines specify the range and number of comparators in a flash converter. In line 4 the required overall yield of, e.g., 95 % is recalculated to the required probability that a pair of adjacent comparators remains monotonic. This number is close to 1. For one bit more and double the number of comparators, this probability is even closer to 1. Now in line 5 a table for a normal distribution is used to find the number of sigmas needed to reach this probability outcome. In line 6 the input referred mismatch is found by dividing the value of one bit by this number of sigmas. Another division by $\sqrt{2}$ accounts for the step from a difference between two

Table 8.1 Comparison of N and N+1 bit flash analog-to-digital converters

		N bit	N+1 bit
1	Input range	V_{ref}	V_{ref}
2	LSB size	$2^{-N}V_{ref}$	$2^{-(N+1)}V_{ref}$
3	Number of comparators	$2^N - 1$	$2^{(N+1)} - 1$
4	Probability per comparator pair for 95 % ADC yield	$p_N = \sqrt[2^N-1]{0.95}$	$p_{N+1} = \sqrt[2^{(N+1)}-1]{0.95} \approx \sqrt{p_N}$
5	Excess factor σ's in $N(0,\sigma)$	$z_N \approx 3\ldots4$	$z_{N+1} \approx z_N + 0.3$
6	Input referred random error	$2^{-N}V_{ref}/z_N\sqrt{2}$	$2^{-(N+1)}V_{ref}/z_{N+1}\sqrt{2}$
7	MOS pairs in comparator	3	3
8	Random error per pair	$\sigma_N = 2^{-N}V_{ref}/z_N\sqrt{6}$	$\sigma_{N+1} = 2^{-(N+1)}V_{ref}/z_{N+1}\sqrt{6}$
9	Area per MOS	$WL = A_{VT}^2/\sigma_N^2$	$WL = A_{VT}^2/\sigma_{N+1}^2$
10	Capacitance of all gates	$3 \times 2^N A_{VT}^2 C_{ox}/\sigma_N^2 = 18S_N^2 2^{3N} A_{VT}^2 C_{ox}/V_{ref}^2$	$3 \times 2^{N+1} A_{VT}^2 C_{ox}/\sigma_{N+1}^2 = 18S_{N+1}^2 2^{3(N+1)} A_{VT}^2 C_{ox}/V_{ref}^2$

comparators to a single comparator. In lines 7 and 8 this mismatch budget is divided over 3 relevant pairs of transistors in a comparator. In line 9 the required gate area is shown and finally line 10 gives the total capacitance per analog-to-digital converter. The input capacitance is dominated by the resolution 2^{3N}, yet it remains important to come to a high ratio between reference voltage (equals the signal swing) and mismatch coefficient. The difference between the input capacitance of an N bit and an $N+1$ bit converter is

$$\frac{C_{in,.N+1}}{C_{in,N}} = \frac{8S_{N+1}^2}{S_N^2} \approx 10 \quad \text{for } N = 5,\ldots,8 \tag{8.22}$$

The last lines in Table 8.1 show that the technology factor $A_{VT}^2 C_{ox}$ with the dimension of energy (Joule) is determining the outcome, compare Fig. 5.48. This figure gives a first indication that the performance in 28-nm high-k factor metal-gate processes can be significantly improved.

Example 8.5. Compare the input capacitance of a 7-bit and an 8-bit flash converter with 1-V input range in a 65-nm process.

Solution. Table 8.2 shows the effect of increasing the resolution with one bit in a flash analog-to-digital converter.

Example 8.6. In a simple flash converter the size of the transistor input pair of the comparators is 20/5 µm in a process with $A_{VT} = 15\,\text{mV}\mu\text{m}$. The expected input signal is $1\,V_{peak-peak}$. What resolution limit (if monotonicity is required with 99 % yield) do you expect?

Solution. Monotonicity means that in case of an increasing input voltage, the comparator connected to a lower reference voltage switches before the comparator

Table 8.2 Comparison of 7- and 8-bit flash analog-to-digital converters

		7 bit	8 bit
1	Input range	1 V	1 V
2	LSB size	7.8 mV	3.9 mV
3	Number of comparators	127	255
4	Probability on monotonicity per comparator pair for 95 % ADC yield	0.99960	0.99980
5	Excess factor σ's in $N(0,\sigma)$	3.35	3.55
6	Allowed input referred random error	2.33 mV	1.10 mV
7	Number of MOS pairs in comparator	3	3
8	Random error per input pair	1.34 mV	0.63 mV
9	Area per MOS $A_{VT} = 3.5\,\text{mV}\mu\text{m}$	6.8 μm^2	30.7 μm^2
10	Capacitance per MOS $C_{ox} = 12.6\,\text{fF}/\mu\text{m}^2$	86 fF	387 fF
11	Capacitance of all input gates	11 pF	99 pF

Table 8.3 The calculated yield for 6, 7, and 8 bit

N	V_{LSB}	z	$P(x)$	$P^{2^N}(x)$
6	15.6 mV	7.1	$1–10^{-11}$	0.999999
7	7.8 mV	3.55	1–0.0002	0.975
8	3.9 mV	1.77	1–0.0384	4×10^{-5}

6-bit is producible, 7-bit is publishable, 8-bit is theoretical

connected to a V_{LSB} higher reference voltage. The standard deviation of the input pair and trip level is calculated as: $\sigma_{trip,i} = A_{VT}/\sqrt{WL} = 1.5\,\text{mV}$, ignoring other contributions. The nominal difference between two trip levels is $V_{LSB} = 2\text{V}/2^N$. As both trip levels suffer from uncertainty, the standard deviation of the difference equals $\sigma_{i+1,i} = \sqrt{\sigma^2_{trip,i+1} + \sigma^2_{trip,i}} = 1.5\sqrt{2}\,\text{mV}$. The probability $P(x)$ that the difference between two trip levels stays within V_{LSB} corresponds to the value of a normal $N(0,1)$ distribution for the quantity $z > \sigma_{i+1,i}/V_{LSB}$.

Monotonicity in a converter requires that all 2^N differences are within one LSB. So this probability requires to multiply all 2^N individual probabilities. Table 8.3 shows that a 7-bit converter is at the edge of acceptable yield.

The faster way to this result uses Fig. 8.32.

8.2.3 Decoder: Thermometer

In many analog-to-digital converters the decoding of the comparator decisions into a digital output word is not a big challenge, though the algorithm behind the decoding is sometimes interesting. Figure 8.34 shows a basic wired NOR decoding scheme for a flash converter.

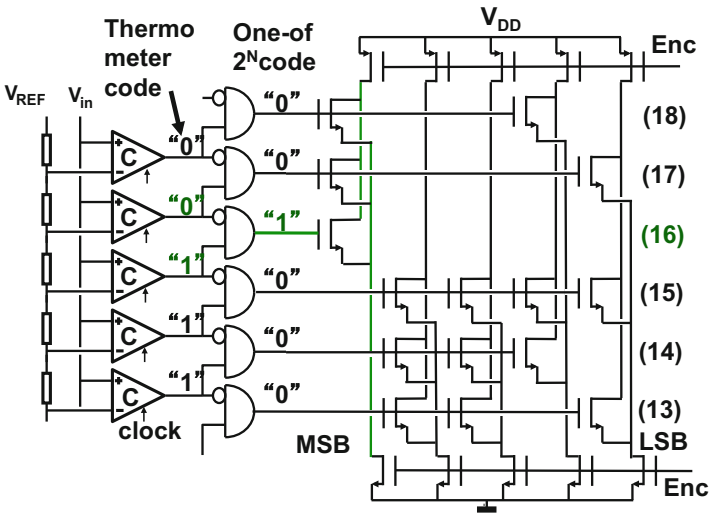


Fig. 8.34 A wired NOR-based decoder scheme

The decoding starts by a simple gate that converts the thermometer code at the output of the comparators in a 1-of-2^N code. This is the digital form of the mathematical derivative function. The function will indicate the pair of comparators where the input signal is between the reference values. The corresponding decode line is connected to a matrix of transistors laid-out in a straight binary code. The upper PMOS transistors charge the vertically running lines to V_{DD}. Shortly after the comparators have taken their decisions, the encode signal "Enc" is activated. The PMOS devices are turned off and the lower NMOS transistors are turned on. The previously charged lines can only be discharged if the matrix transistors are conductive via the 1-of-2^N code.

The above sketched operation can be disturbed by various mechanisms. If mismatch creates a situation where a lower or higher comparator switches too, more than one wired-NOR input line will be active and more decoding lines are discharged. These errors are called "sparkles" or "bubbles." This situation can also occur in the presence of high slew rate signals or meta-stability errors. Most of these sparkles will create a deviating code, however, if the sparkle affects a decode region around a major bit, full-scale errors can be the result. There are numerous ways to avoid that these sparkles upset the decoder. Figure 8.35 simply extends the thermometer decoding with an additional input. Many other solutions exist where different trade-offs with speed are made.

In medium performance applications this scheme will work fine. However if the operating frequency is increased to the limits of the technology, performance problems arise. At high operating frequencies the inherent capacitive load of a wired NOR structure becomes a burden. Buffering is required at the cost of power. In advanced processes the parasitic wiring capacitance fortunately reduces significantly thereby allowing fast processing of the decoding signals.

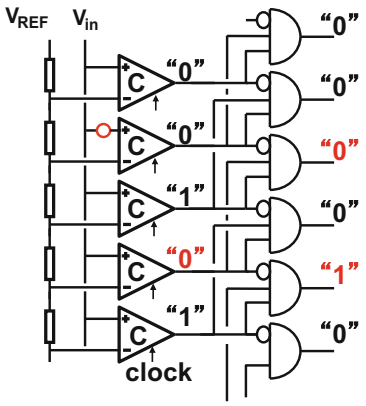


Fig. 8.35 A wired NOR-based decoder scheme with bubble correction

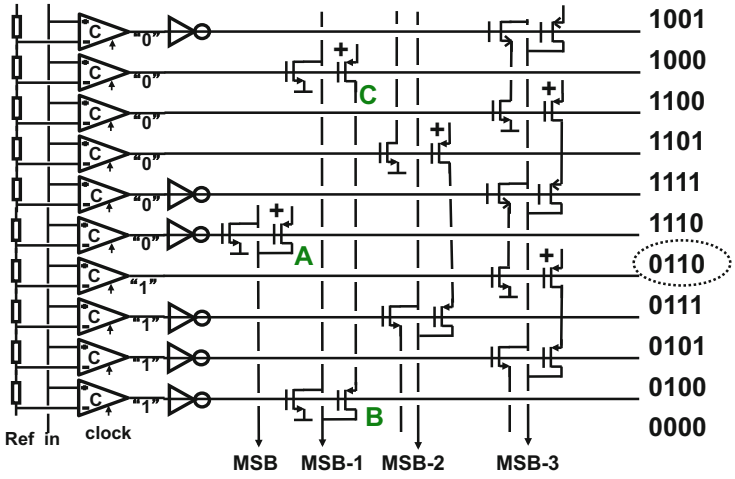


Fig. 8.36 A wired Gray-decoder scheme

8.2.4 Decoder: Gray-Code

The above decoding schemes show a disadvantage in case of meta-stability. A meta-stable comparator output in Fig. 8.34 will affect two decode gates and their associated decode lines. If the meta-stable condition continues, opposing signals may appear in the two decode gates and a major decoding error will happen.

The MSB line in the Gray-decoding scheme of Fig. 8.36 is controlled via a comparator and two inverters. The last inverter, labeled "A," drives the output line. The MSB-1 uses two gates "B" and "C." For a signal on the input terminal lower than the reference voltage of this scheme, all comparators will signal a logical zero. Gate "B" receives a logical one and the NMOS will pull the line to a logical zero state. After the input signal increases to the level corresponding to the encircled

code in Fig. 8.36 the input of gate "B" will go to logical zero just as gate "C." Via both PMOS devices the MSB-1 line is set at a logical one state. If the input signal exceeds the reference levels in this drawing, the input to gate "C" will be a logical one, pulling the MSB-1 line to zero.

Each comparator in a Gray decoder sets one single output line. A potential meta-stable condition does not spread out over more than one cell and can still be resolved if more time is allowed. Half of the meta-stability will happen in LSB-switching comparators, where the resulting consequences are in first instance low. The Gray-code does not solve the metastability problem, but limits the effects.

After the Gray-code a well-known EXOR scheme can translate the data into a straight binary format. The Gray-code is an often used strategy in the first part of the decoder. For large decoders the remaining decoding can be done in conventional style.

8.2.5 *Decoder: Wallace*

A rigorous solution to the decoding problem is to add up all comparator outputs, Fig. 8.37. A structure performing that task is a Wallace tree consisting of a large number of full-adder cells. Every full-adder cell takes three signals of equal weight 2^k and combines that into two signals of weights 2^k and 2^{k+1}. It is functionally not important which wires are connected to which adder. Practically a designer will strive to minimize the number of layers. Every full-adder reduces the signal count by one. So reducing $2^N - 1$ comparator outputs to N lines takes $2^N - 1 - N$ full-

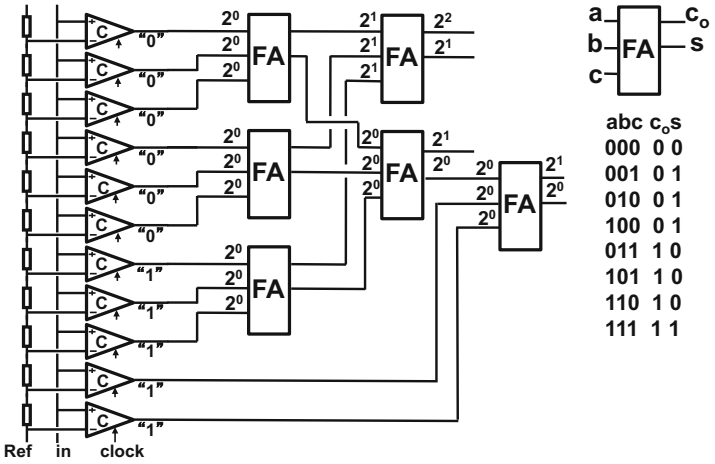


Fig. 8.37 In a Wallace tree full adders combine three digital signals of equal weight into two signals

adders. This decoding solution eliminates the risk of non-monotonicity as every comparator that switched on is counted, irrespective of the order in the ladder. This feature is beneficial in some circuit topologies. A Wallace tree generates quite some digital activity. Especially when the layers are not very uniformly filled, some adders will first first go high and picoseconds later low, consuming a full CV^2 packet of energy. Also a meta-stable condition can spread out easily. With improving digital performance, however, more designers turn to this decoder.

8.2.6 Averaging and Interpolation

The main problem for improving the accuracy of a flash converter is the mismatch of the comparator stages. This mismatch directly translates in INL and DNL performance loss. In order to alleviate this problem the ratio between signal amplitude and input referred random offset of the comparators must be improved. Since the early nineties two techniques are applied for this purpose: averaging and gain stages with interpolation. The random mismatch problem in traditional flash converters is based on the ratio of the input signal and random offset in one comparator stage. The fundamental observation in [198] is to combine multiple input stages. The signals of these stages are added up linearly, while their mismatch adds up in a root-mean-square manner. Consequently their ratio will improve if more input stages are combined. Figure 8.38 shows the topology applied to a simple comparator topology. The input stages of the comparators generate a differential current that form a differential voltage over the resistors R_t. These currents are combined over various stages through coupling resistors R_c.

Figure 8.39 (left) shows a part of the resistor network. R_e is the equivalent impedance seen to the left and R_t and R_c add another stage to this network. Now the equivalent resistance after the addition of these two elements should again be equivalent to R_e. Some arithmetic gives [198, 199]:

$$\frac{R_e}{R_t} = -\frac{1}{2}\alpha + \frac{1}{2}\sqrt{\alpha^2 + 4\alpha} \tag{8.23}$$

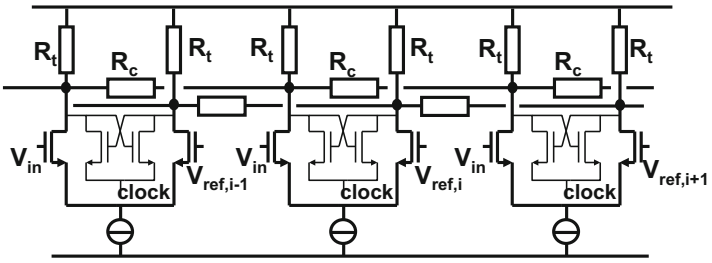


Fig. 8.38 The averaging scheme after [198]

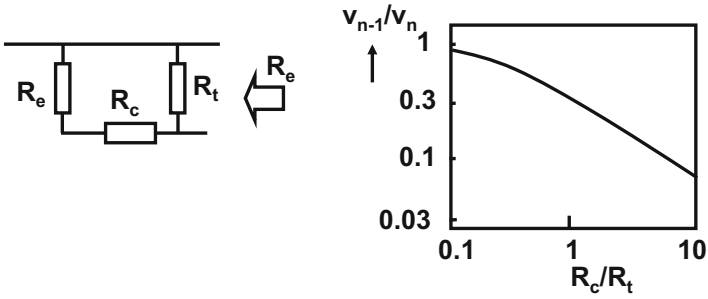


Fig. 8.39 Starting point for the analysis is the equivalent resistor network on the *left*. Equation 8.24 is shown to the *right* [199]

where $\alpha = R_c/R_t$. Based on this equivalent impedance the effect of the coupling resistor R_c on various performance aspects can be derived. If one input pair creates an offset current resulting in a voltage v_n over resistor R_t, then the impact of v_n on the neighboring voltage v_{n-1} is

$$\frac{v_{n-1}}{v_n} = \frac{-\alpha + \sqrt{\alpha^2 + 4\alpha}}{\alpha + \sqrt{\alpha^2 + 4\alpha}} \tag{8.24}$$

With $\alpha = 1$ this factor is 0.38. Equation 8.24 is depicted in Fig. 8.39 (right). A small α (low R_c) increases the coupling between the node voltages and improves the ratio between the linear signal component and the stochastic variation. Also the range of comparators that contribute is increased. This range cannot be made infinitely long because only comparators with input stages operating in the linear regime can effectively contribute. As an example [200] used an averaging over five stages.

An issue with the averaging technique is the termination of the range. A low α factor means a low coupling resistor value. A larger number of neighboring comparator stages are combined and a better random offset reduction is achieved. Near the limits of the range this means that a relatively large number of additional comparators are needed for achieving also an offset reduction at the extremes of the converter range. Scholtens and Vertregt [199] proposes to use a dedicated termination cell at the end of each side of the comparator structure, which reduces this problem. Other techniques involve folding or are based on the Möbius band.

A second form of improving the ratio between the signal and random mismatch is shown in Fig. 8.40. In this scheme a group of additional amplifiers is placed before the comparators. The amplifiers locally boost the difference voltage between the signal and the adjacent reference voltages by a factor 2 to 4. The result is applied to an interpolation ladder with comparators. A typical gain stage will serve 4–8 comparators. Now the signal is amplified by the gain section with adjacent reference voltages. The other sections rely on their much larger overdrive. Obviously the mismatch problem in the comparators is reduced by the amplification factor, while

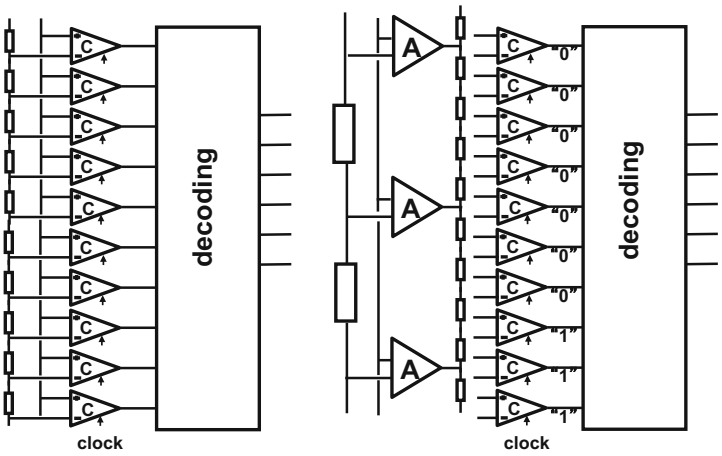


Fig. 8.40 Gain stages (A) are applied to increase the voltage swing on the comparators

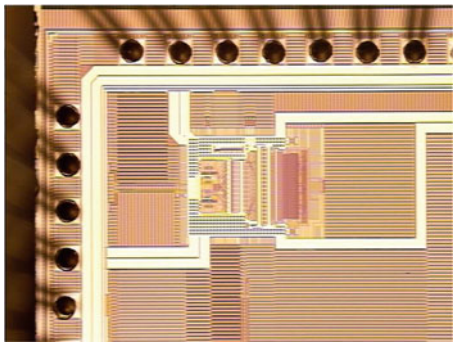


Fig. 8.41 A picture of a 6-bit flash ADC converter [199]

the mismatch in the amplifiers mostly affects the INL and can be reduced as there are far less amplifiers than comparators. The trade-off in this scheme is between the additional power needed for the high-performance gain stages and the reduction on the side of the comparators.

Figure 8.41 shows a flash converter placed in a system-on-chip design. Frequently averaging and interpolation are combined into the same design [199, 201].

8.2.7 Frequency Dependent Mismatch

So far, no frequency dependencies of the input gain stages have been considered. Figure 8.42 shows an example of a comparator where the offset reduction is accomplished by means of a gain stage before the latch stage. These elements can be identified in most comparators. The latch stage is considered ideal except for a load

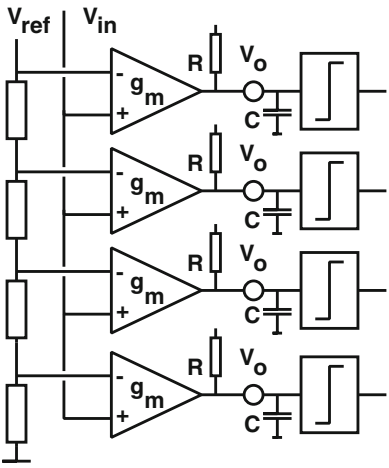


Fig. 8.42 Random offset reduction by means of a gain stage

capacitor C and a random offset source V_o. Both are related to the latch transistor dimensions by $C = WLC_{ox}$ and $\sigma_{Vo} = A_{VT}\sqrt{N_T/WL}$. A_{VT} is the process constant for threshold matching (Sect. 5.3) and N_T is the number of latch transistor pairs that contribute to the random offset. For low frequency operation the input referred mismatch due to the latch mismatch is calculated by dividing the latch mismatch by the effective DC-gain. For high-frequency operation the input-referred random offset due to the latch is given by σ:

$$\begin{aligned}\sigma &= \frac{\sigma_{Vo}(1+j\omega RC)}{g_m R} \\ &\approx \frac{A_{VT}\omega C_{ox}\sqrt{N_T WL}}{g_m} \qquad \omega RC > 1 \end{aligned} \tag{8.25}$$

This is only a rough approximation, which must be adapted for the sample-and-hold operation, timing and different architectures. It indicates that random-offset reduction is frequency limited and depends on design (N_T, g_m), technology (A_{VT}, W, L), and power (g_m, number of comparators). Example: in a 1 µm CMOS technology $\sigma_{Vo} \approx 20\,\text{mV}$, $gm/2\pi C \approx 250\,\text{MHz}$, the contribution to the input referred random offset is 0.8 mV at 10 MHz bandwidth and increases linearly with the input frequency.

8.2.8 Technology Scaling for Flash Converters

In 90-nm, 65-nm, and 45-nm various analog design constraints appear that affect the performance of converters. A few of these effects are

- The power supply drops to 1 V. As a consequence all signal levels will be lower. Differential design is imperative, yet it will be difficult to handle more than 0.3 V of single sided signal swing.
- Thresholds go down in order to keep the current drive at an acceptable level. Low threshold voltages are not convenient for cascode structures, because the input and output DC-levels start to differ too much..
- The mismatch for a fixed area device is less predictable due to additional implants. At 45 nm the same transistor size may show comparable random variation as in 90 nm.
- The beta factor is low for minimum size devices, it can be improved by backing-off on the designed gate length.
- The construction of the transistor with high additional pocket dopes near channel to control the electric fields (halo implant), leads to a low intrinsic amplification factor of the device.
- Deep n-well constructions reduce substrate noise and allow floating NMOS devices next to the floating PMOS devices.
- Gate leakage of transistors with less than 1.5-nm effective gate-oxide becomes an issue as the gate current (1–10 nA/μm gate width for 1.2 nm gate oxide) will load the ladder.

8.2.9 Folding Converter

A variant of a flash converter is the folding analog-to-digital converter [2, 181, 202, 203] in which a pre-processing stage "folds" the input signal. Figure 8.43 shows the general idea: the input signal is folded into (in this example) 8 sections. The resulting folded range is applied to the succeeding analog-to-digital converter and

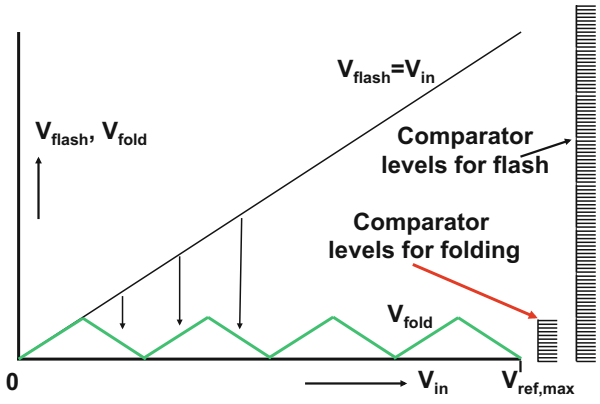


Fig. 8.43 A folding analog-to-digital converter folds the input signal into a smaller signal range

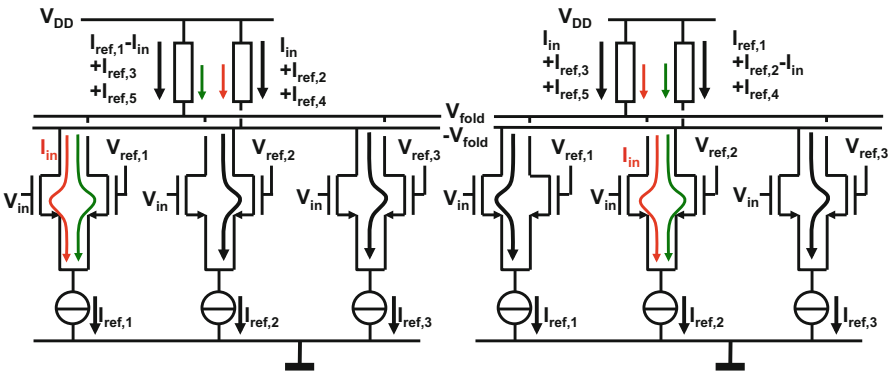


Fig. 8.44 The basic folding circuit: the input pairs are designed to reverse the signal current in the top resistors for each folding section. *Left*: the signal is in the operation range of the first differential pair. *Right*: the signal has increased and uses the second differential pair. The first and third differential pairs are saturated

is reduced to 1/8 of the original range. Next to this analog-to-digital converter, this method requires a circuit that performs the folding operation and a coarse analog-to-digital converter to keep track of the section number. This partitioning reduces the total number of comparators. If an 8-bit flash converter is split in, e.g., F=8 folding sections (3-bit) and a remaining 5-bit fine flash converter, a total of $2^3 + 2^5 = 40$ comparators are needed.

The folding principle was originally developed in bipolar technology [202, 204]. Later on the folding principle was applied in CMOS analog-to-digital conversion [181, 183]. In both technologies the basic topology is a parallel arrangement of differential pairs. These pairs are driven by the input signal and have each a local reference voltage of 1/2F, 3/2F, 5/2F,... of the full-scale reference voltage. In Fig. 8.44 the first three sections of a folding stage are depicted. In this example the reference voltages are chosen for $F = 8$ folding sections spaced at $1/F$ fractions of the overall reference voltage. The transconductance of each stage is designed to cover an input voltage range of $\pm 1/2F$ of the reference voltage. In this way the F stages cover the entire input range. The idea is that the signal itself will select the input stage. The output of the folding stage on an input ramp is a triangle shaped signal with a periodicity of half of the number of fold sections. A full-range sinusoidal input signals will be multiplied with this triangular function and generate frequency folding signals at $(F+1)$ and $(F-1)$ times its own frequency. A full-range input signal generates a folding signal with $(F + 1)$-th and $(F - 1)$-th harmonics. Smaller signals use less folding stages and result in lower frequency harmonics. This multiplication effect is a potential cause for distortion products in a folding analog-to-digital converter when these higher order signals proceed to the output in case of asymmetries and mismatches. Moreover the bandwidth of the output stage needs to accommodate these frequencies. After the folding stage some gain can help to reduce the accuracy requirements for the comparators in the succeeding flash stage.

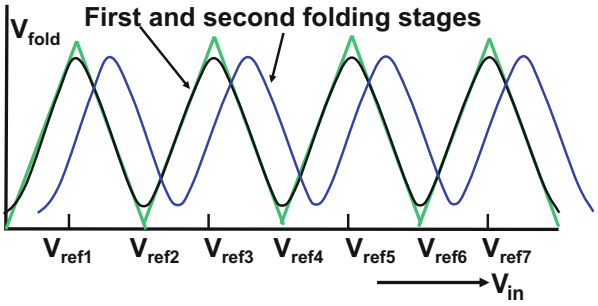


Fig. 8.45 The basic folding circuit suffers from distortions near the switch points. In a practical design two folding circuits with a mutual offset allow an interpolation scheme to determine the switching levels

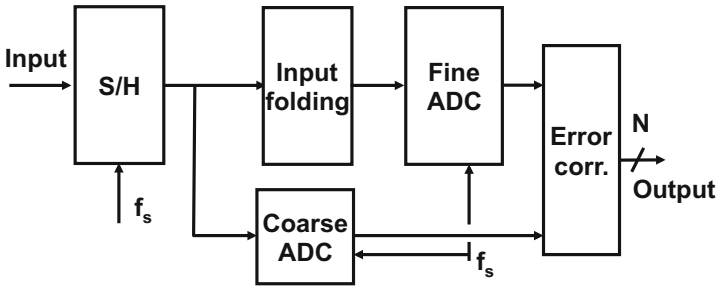


Fig. 8.46 A folding analog-to-digital converter architecture incorporating a sample-and hold [181]

The cross-over point between the stages in the basic concept is a weak point in the architecture of Fig. 8.44. In order to mitigate the problems with the cross-over points, mostly a dual folding architecture as in Fig. 8.45 is used. The second folding stage is shifted by half of the range of a single folding section $1/2F$, thereby creating a linear transfer at the cross-over points of the first folding stage. The coarse sub analog-to-digital converter selects the linear part of the transfer curves. More recent designs use a resistive interpolation technique between the outputs of these voltage-shifted folding stages. This results in bundles of transfer curves where the comparators trigger at the cross-over of the appropriate combination of (interpolated) folding signals [181, 200]. Proper design of the pre-processing stage where high-speed and high yield are combined is the critical issue.

The architecture of Fig. 8.46 uses a coarse analog-to-digital converter to detect the folding section. The design requires a sample-and-hold circuit. The main reason is that the input signal uses two paths through this converter: the path via the coarse converter and the path through the folding stage. The outcome of these paths must remain synchronous within one clock period. The advantage of using an S/H circuit is that signal propagation errors in the analog pre-processing are reduced and that architectures can be used that make multiple use of the input signal.

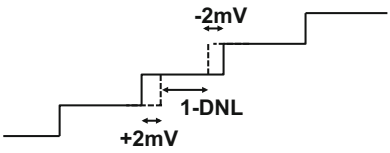


Fig. 8.47 The shift of two trip levels in a worst case situation

The second reason for using a sample-and hold lies in the observation that folding stages generate frequency multiples of the input. A sample-and-hold will reduce this problem to a settling issue. Design experience has shown that a high-speed S&H running at full-signal and bandwidth requires 30 % of the total analog-to-digital converter power budget.

The remaining problem with folding analog-to-digital converters is random offset in the folding input pairs. Any deviation in one of the folding stages will translate in performance loss. An attempt has been made to address this issue [200] by using an interpolation method at the cost of a lot of power. In bipolar technology a performance level of 10 bit has been reported [205], while in CMOS an 8–9 bit level is state-of-the art [181, 183, 200]. A form of calibration can further improve the resolution [206].

Example 8.7. In an AD converter with an input range of 1.024 V the comparator can have an input referred error of maximum/minimum $+2/-2$ mV. What is the best resolution a flash converter can reach if a DNL of maximum 0.5 LSB is required? Which converter type could reach a better resolution?

Solution. A DNL error is caused because the trip levels shift due to comparator mismatch, see Fig. 8.47. In this case a situation can arise where the i-th trip level shifts $+2$ mV, while the $i+1$-th trip level shifts -2 mV. These two errors together may cause a maximum DNL error of 0.5 LSB. Consequently $0.5V_{LSB} = 2 + 2$ mV. An LSB size of 8 mV then allows a 7-bit converter.

Any converter based on a single comparator can perform better: e.g., a successive approximation converter.

8.2.10 Digital Output Power

The result of an analog-to-digital conversion is a digital word changing its value at the speed of the sample clock. Obviously this interface to the digital processing consumes power as is given by the standard digital power equation. Charging a capacitance C_{load} requires $C_{load}V_{DD}^2$ energy from the power supply. Assuming that at every sample pulse half of the N output pins will change state. Only half of those will require energy from the power supply while the other half switching outputs discharge into ground. This yields

$$P_{dig-out} = Nf_sC_{load}V_{DD}^2/4$$

With $f_s = 1$ GHz, $V_{DD} = 1$ V, $N = 10$ bit and on-chip capacitive load of $C_{load} = 1$ pF the power consumption is 10 mWatt. Much of this power contains signal related components that can cause serious performance degradation if these components influence the analog signals. The digital output power is preferably delivered by a separate power supply.

If an external load must be driven by the digital output, the capacitive load is at least an order of magnitude higher: 10–20 pF. The power needed in the digital output driver climbs to 0.2 W or more. Special precautions must be taken to avoid coupling between these drivers and the rest of the analog-to-digital converter and other analog circuits. Digital signals are full of noise and distortion. In case unexplainable distortion appears, always check whether there is a path from the digital output into the converter.[7] Separate power wiring, power bondpads interleaved between the output bits. Also the heat production of the drivers can be of importance.

There are various ways to reduce the effect of the digital output section. Examples are: reducing the output load[8] by connecting an external digital register closely to the analog-to-digital package, low-swing outputs, small damping resistors in series. More expensive measures implement differential output ports or digitally coded outputs. Low sample-rate converters can use a parallel-to-series output. All N data bits are shifted out via a single bondpad at a clock rate of $N \times f_s$. In fact, this will not reduce the needed output power, but the much higher frequencies will contain less energy in the signal bandwidth.

8.2.11 Mismatch Dominated Conversion

In this entire chapter mismatch is regarded as a severe problem for performing analog-to-digital conversion. Nevertheless it is certainly possible to use mismatch to the advantage [207]. The comparators in Fig. 8.48 are designed to have a large input referred mismatch. The group comparators connected to $V_{REF,1}$ will span a voltage range of several tens of mV and follow a Gaussian distribution. In order to increase the range a second group of comparators is connected to $V_{REF,2}$ overlapping the first range. The total network will show a two-peak probability density function. A summation network (e.g., a Wallace tree structure) is used to count the number of flipped comparators. This number allows to estimate the input voltage via the probability density function. The Wallace tree is a very power hungry network and the input range is determined by the A_{VT} coefficient. Several other variants were

[7] A potential test is to lower the digital amplitude by decreasing the digital power supply. If the distortion is reacting, then a coupling path must exist.

[8] An often encountered error in a measurement set-up is a direct connection between the ADC chip and the input port of a laptop. Certainly some spurs related to the internal processing will be visible in the analog-to-digital conversion spectrum.

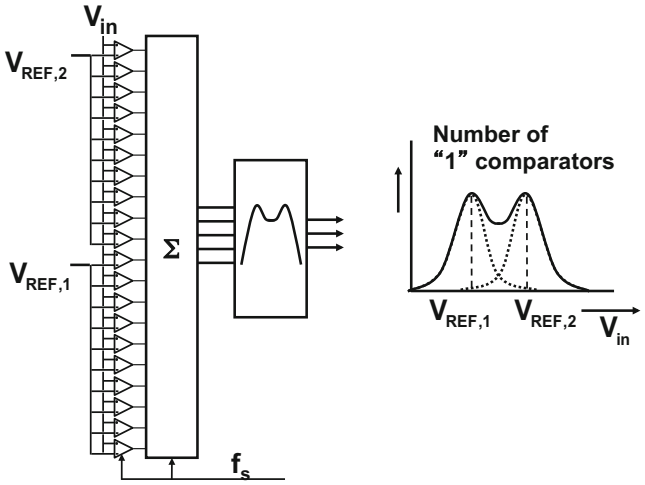


Fig. 8.48 Using the mismatch of the comparators to cover an input range

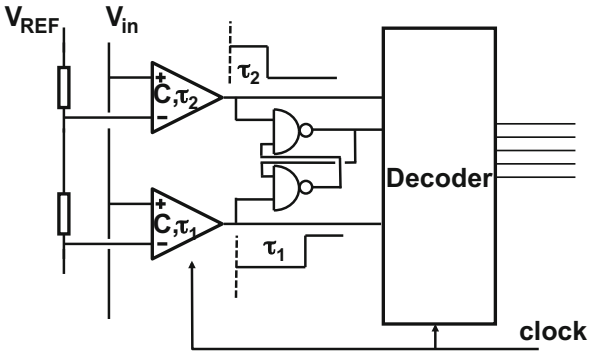


Fig. 8.49 The delay of a comparator is dependent on the difference between input and local reference. For a certain input level the delay of both comparators differs which is measured with a simple set-reset latch. In [208] one bit of additional resolution is claimed

published by the same authors, e.g. to select those comparators that form a (more-or-less) equidistant frame of reference values and disable the unused circuits.

Another form of exploitation of circuit effects is shown in Fig. 8.49 [208]. As indicated in Sect. 8.1.2 the decision delay of a comparator depends on the voltage difference between input and reference. This time delay increases when the voltage difference decreases. The two comparators in Fig. 8.49 show equal delay when the signal is exactly in the middle between both comparator trip levels. And different delays when the input voltage is closer to one or the other reference level. The delay is detected with a set-reset latch and creates one bit of additional resolution. The sensitivity for noise, voltage spikes, power supply variations, etc. must be carefully examined.

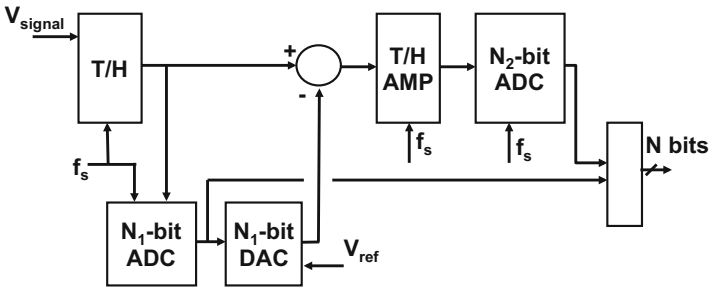


Fig. 8.50 Block diagram of a two-step converter

8.3 Subranging Methods

For accuracies of 8 bits or more flash converters are no economical solution due to the exponential growth of area, power, and input capacitance. Subranging methods allow to achieve higher resolutions at the cost of sampling speed. In Fig. 8.50 the signal flow in a two-step converter is shown. The signal is sampled and held at the input. A limited resolution coarse analog-to-digital converter (e.g., a small flash converter) with a resolution N_1 estimates the signal. This information is fed to a digital-to-analog converter and subtracted from the held output signal of the left T&H circuit. The subtraction results in a residue signal, with a maximum amplitude fraction of 2^{-N_1} of the input range. This signal is amplified in a second T&H gain stage. A second fine converter with a resolution N_2 converts this residue signal. This simple approach results in a converter with a resolution of $N_1 + N_2 = N$. In this elementary approach of this converter only $2^{N_1} + 2^{N_2} \ll 2^N$ comparator circuits are needed. Converters based on this principle are known under names as: "subranging," or "coarse-fine," or "two-step" analog-to-digital converters. A simple way to understand a subrange converter is the analogy with a ruler having a centimeter and millimeter scale. When an object is measured, first the centimeters are read and after that the millimeters.

A rising signal in a subranging converter experiences the transition points of the coarse section. In between the same fine-converter is used for every coarse section in the transfer curve. Therefore the INL curve will show repetitive patterns, as in Fig. 8.51.

In this type of converter additional components are present such as T&H circuits and an additional digital-to-analog converter. In a design where $N_1 + N_2 = N$ the coarse analog-to-digital converter, the digital-to-analog converter, and the subtraction point must operate within an LSB of the overall converters resolution in order to avoid loss of precision.

The speed of the subrange converter is in first instance limited to the time needed for entire operation: the input track-and-hold, the first coarse conversion, the digital-to-analog settling, the subtraction and the second fine conversion. This processing can, however, be pipelined over two sample periods by inserting a second

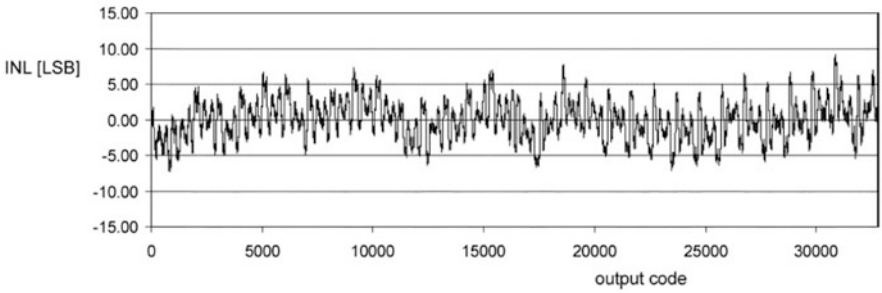


Fig. 8.51 Subranging converters show a repetitive pattern in their transfer curve. This *curve* is from a 15-bit resolution converter [209]

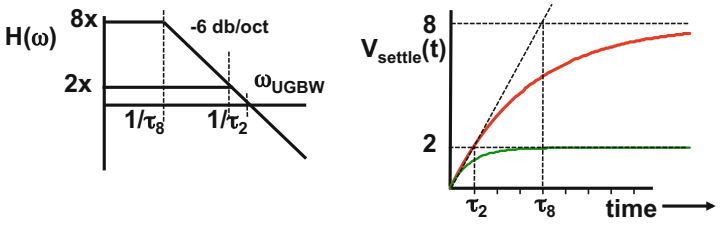


Fig. 8.52 Frequency and time response of a system with fixed unity-gain bandwidth. *Left*: more gain means a proportionally larger time constant. *Right*: the time response on a step function. Again the settling of larger amplification costs settling speed

T&H circuit behind the subtraction. After the subtraction the amplitude of the remaining signal is small. With this T&H circuit some amplification is performed and consequently the effects of errors in the succeeding processing are suppressed. Now two successive samples are processed in a pipelined fashion, and the sampling speed is limited to the processing speed of just a single section.

The amplification between stages is a crucial design decision. Figure 8.52 compares two times amplification with eight times. For a well-designed amplifier (OTA or opamp) the technology and the power determine the capacitance and the current, resulting in a maximum unity-gain bandwidth value (UGBW, ω_{UGBW}). Some variation with opamp topologies is possible, yet in many fast opamps with a first-order dominant pole characterized by -6 dB/octave, the input transconductance and the load or Miller capacitor define the UGBW. Using this amplifier in a $2\times$ or $8\times$ amplification mode results in slower settling for the higher gain. In the frequency domain $2\times$ gain means $\tau_2 = 2/\omega_{UGBW}$ and $8\times$ gain means $\tau_8 = 4 \times \tau_2$.

In the time domain the settling behavior on a step input shows the same properties: the settling of an $8\times$ stage is four times slower than of a $2\times$ stage, where in the frequency domain the UGBW and the first order slope are fixed, in the time domain the $dV/dt = I/C$ near zero is fixed.

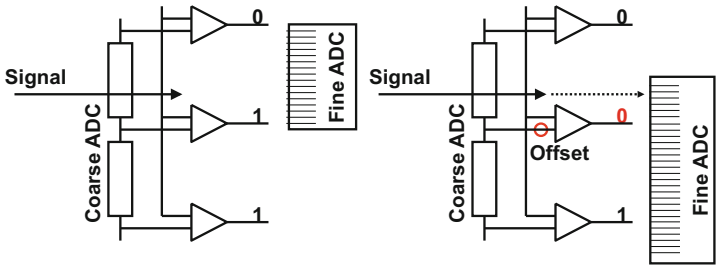


Fig. 8.53 Coarse-fine conversion: on the *left-hand side* the coarse flash converter takes a correct decision and the fine-converter receives a signal within its range. On the *right-hand side* one comparator has offset and the coarse converter indicates the wrong range for the fine converter. As the fine converter has additional levels, still the out-of-range signal can be correctly converted. The succeeding digital logic will correct the wrong coarse decision

Therefore a one-bit-pipeline stage in the same technology and power can be significantly faster than a subrange converter with, e.g., 8× amplification. The disadvantage is of course that succeeding errors in a one-bit converter are less suppressed.

8.3.1 Overrange

A disadvantage of the set-up in Fig. 8.50 with $N_1 + N_2 = N$ is the need for a perfect match between the ranges of the first and second converters. If the first (coarse) converter with random offset decides for the wrong range for performing the fine conversion, there is no correct conversion possible. By adding additional levels on both sides of the second converter range, see Fig. 8.53, errors from the first converter can be corrected [210]. Figure 8.54 shows the span of ranges in case the intermediate amplification is 2×. This "over-range" limits the accuracy requirements on the coarse analog-to-digital converter. In some designs the overrange doubles the total range of the fine converter.

Full accuracy (N-bit level) is still needed in the digital-to-analog converter and in the subtraction circuit. The coarse converter can now be designed to a much more relaxed specification. Implicit in this method is that the residue portion of the signal achieves the accuracy requirements of the N-bit conversion.

The sample rate of subrange converters is limited by settling mechanisms, specifically of the input T&H and the amplifying second pipeline stage. The overrange feature allows to reduce the time allocated for the signal to settle in the chain from coarse converter, via digital-to-analog converter and subtraction node. In Fig. 8.55 (left) the coarse decision has to wait till the T&H output is settled. Due to the overrange, the coarse decision can be scheduled at an earlier moment in the timing, see Fig. 8.55 (right). The resulting error and the potential offsets should

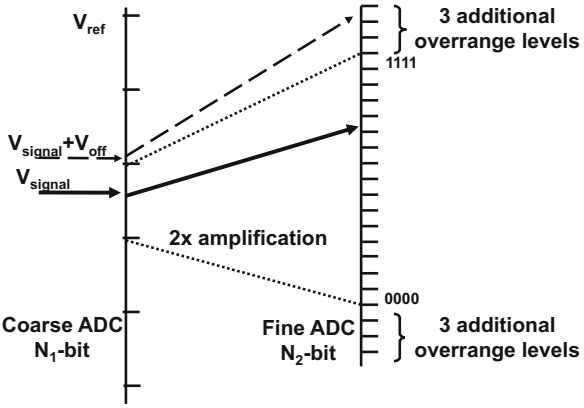


Fig. 8.54 Coarse-fine conversion: on the *left-hand* scale the trip points of the first (coarse) converter are indicated. The remaining signal (*bold line*) is amplified and converted by the second (fine) converter. The *dashed arrow* indicates a situation where the coarse converter has decided for the wrong range. Overrange in the fine converter corrects this decision

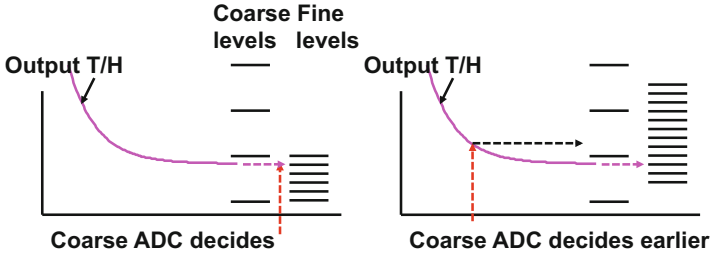


Fig. 8.55 *Left*: timing without overrange, *Right*: as the remaining settling error can be handled by the overrange feature, the coarse decision is shifted forward in time

remain within the overrange section. Of course the fine converter has to wait for a fully settled signal, both from the first and second T&H circuits. This feature of overrange allows a considerable increase of the sampling speed.

The principle of coarse-fine conversion can be extended to three or more stages and resolutions of 14–15 bits, e.g. [209], Fig. 8.51.

8.3.2 Monkey-Switching

In the previous paragraph it was assumed that there is a perfect connection between the coarse digital-to-analog conversion levels defining the coarse transition points and the fine conversion range. The principle of subranging is limited by the quality of the correspondence between these two ranges. Using the earlier analogy of a

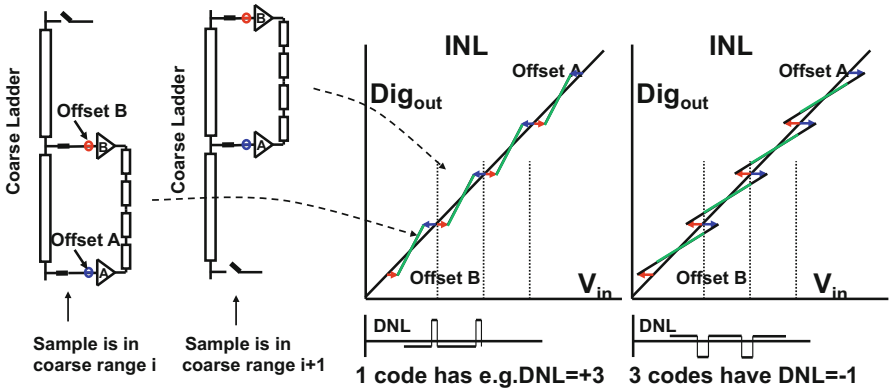


Fig. 8.56 The effect of offset between coarse transitions and the fine conversion in a subranging scheme. INL and DNL are degraded at the coarse code transitions

ruler: the millimeters must exactly fit within the centimeters. There are various ways to link the transitions of the coarse conversion to the fine range. In Fig. 8.56 an elementary connection topology is depicted. The fine ladder is connected to the selected range of the coarse ladder via two buffers A and B. In real converter designs these buffer stages are not always explicit, but can be part of the signal processing. These buffers can have a gain of 1 or more. If these buffers suffer from offsets, the transferred range defined by the coarse ladder will be stretched or shrunk when it arrived at the fine ladder. In Fig. 8.56 the buffers A and B move over the coarse ladder in a fixed mutual position: A is connected to the low-side of the coarse-ladder tap and B to the high side and this order remains if, e.g., the signal increases and the sample is located in the adjacent coarse-ladder segment. This sequence causes that the value of a specific tap on the coarse ladder is passed on to the fine range by adding either the offset of buffer A or the offset of buffer B. If both offsets differ, there will be an INL and a DNL error at the transition points which shows up in the integral and in the differential linearity plots. Especially the sharp transition at the coarse conversion transitions leads to large DNL errors.

In Fig. 8.57 the control of the connections is different: if the converter decides to connect to a higher segment, the lower buffer (A in this example) is disconnected and reconnected to the top side of the next higher segment, while buffer B remains connected to the same tap of the coarse ladder. When increasing the input voltage slowly the buffers will alternately connect to the taps of the coarse ladder. This method is often referred to as "monkey-switching".[9] The DNL transitions are strongly reduced. However, the INL errors remain and are visible together with the errors in the fine converter as a repetitive pattern, see Fig. 4.13 (lower, left).

[9]Some similarity exists with the way a monkey climbs a tree.

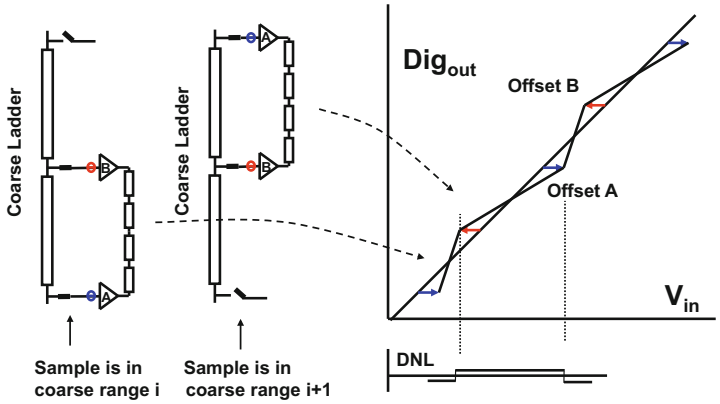


Fig. 8.57 The effect of offset can be reduced by appropriate switching schemes. Only the INL is seriously degraded at the coarse code transitions while the DNL is reduced

Obviously the generated fine converter code must be inverted for the cases where the fine converter is flipped-around, to obtain the correct value.

In [211] the coarse and fine converters are connected via T&H buffers. This subranging architecture with a 7-bit coarse and 6.6-bit fine conversion results in a nominal resolution of 12 bit. A performance is reached of 10.1 ENOB at Nyquist frequency with a sampling rate of 40 Ms/s and 30 mW power.

The setup of the timing is essential especially if additional time periods are allocated to offset cancellation. Typically the coarse converter is activated after some 10–20 % of the hold period. This allows maximum time for the digital-to-analog converter, the subtraction mechanism and the succeeding amplifier to settle.

An alternative to the standard subrange scheme avoids the subtraction and the issues due to offsets between fine and coarse sections. A resistive ladder structure feeds both the coarse and the fine converter sections [139, 190, 192, 212, 213] as shown in Fig. 8.58. The coarse comparators are connected to ladder taps spaced at 2^{N_1} LSB positions apart. The decision of the coarse converter will select a row of switches which is then fed to the fine comparators. Another extension [190, 192] is to use two banks of fine comparators that will alternately digitize the signal, thereby allowing more time for the settling process. Haas et al. [192] also applies overrange. In [212] a capacitive interpolator is used to form the intermediate values for the fine conversion.

These methods use the main resistive ladder structure for both the coarse and fine conversion. This requires special measures to keep the spurious voltage excursions on the ladder under control. In [139] this is realized via a second low-ohmic ladder in parallel to the main ladder similar to Fig. 7.20. Typical effective resolutions are in the range of 8 bit at 40 Ms/s.

A radical proposal for coarse-fine analog-to-digital converter is shown in Fig. 8.59 [214]. The input quantizer is a 6-bit successive approximation analog-to-digital converter that reduces the input range significantly. The 6-bit input

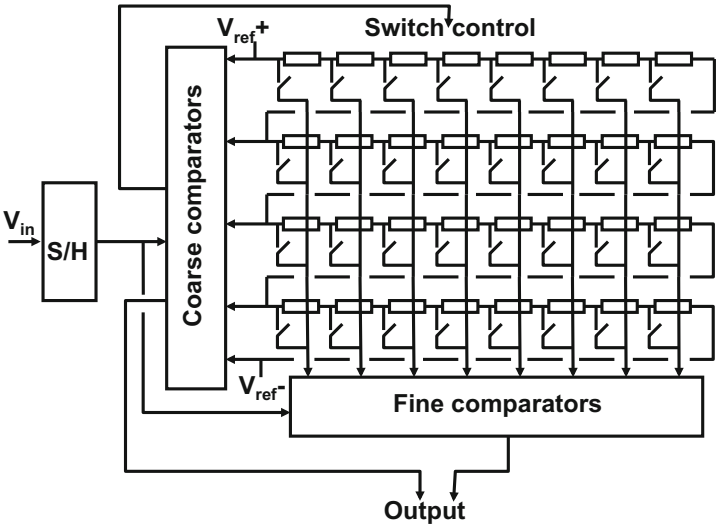


Fig. 8.58 A subranging analog-to-digital converter based on a switched resistor array [139, 190, 212]. For ease of understanding the resolution in this plot is limited to $N_1 = 2, N_2 = 3$

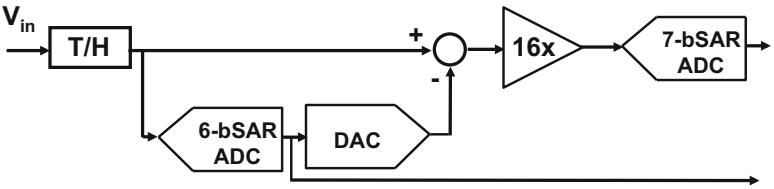


Fig. 8.59 A 6-bit successive approximation analog-to-digital converter strongly reduces the input range [214]

quantization would reduce to residue to a level where $2^6\times$ amplification is possible. However, by reducing the gain to $16\times$ the linearity requirements of the summation and multiply operation are easier met at the cost of the loss of timing for the successive approximation operation, compare also [297]. The successive approximation converters can be designed with self-timed logic thereby speeding up the design. Lee and Flynn [214] reaches a SINAD of 64.4 dB (10.4 ENOB) for a 20 MHz signal sampled at 50 Ms/s for 3.5 mW.

The dominant implementation of stand-alone subrange analog-to-digital converters for industrial applications[10] uses a three-stage approach [215], see Fig. 8.60. The first stage typically converts 4–5 bits and is equipped with a high-performance

[10] See data sheets from ADI, TI, and NXP. Some product numbers: AD9640, AD6645, ADS5474, and ADC1410. In some data sheets these converters are called pipeline converters. In the terminology of this book they are classified as subrange converters.

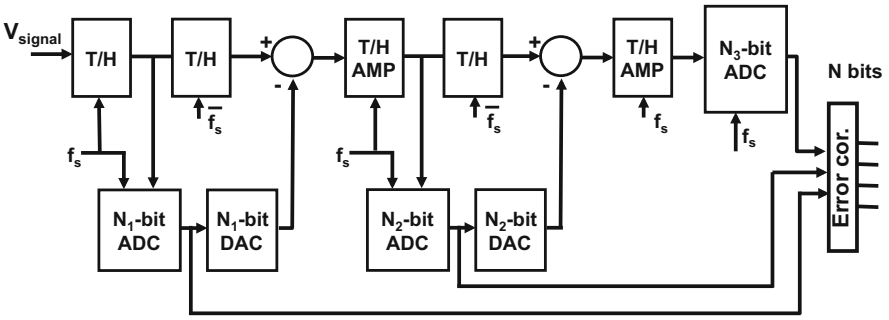


Fig. 8.60 A subranging analog-to-digital converter with three sections. A typical partitioning uses $N_1 = N_2 = 5$ bits and $N_3 = 4$ bit for a 14-bit converter

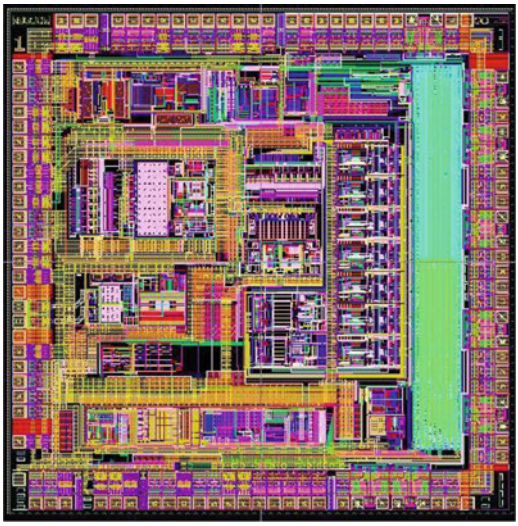


Fig. 8.61 Lay-out of a 16-bit 125 Ms/s coarse-fine analog-to-digital converter ENOB = 11.6, 570 mW, 600 MHz input BW, LVDS outputs, INL = ±4, DNL = ±0.95. Courtesy Ph. Gandy, NXP Caen

track-and-hold. The reduction of the distortion in the track-and-hold is the dominant challenge in these converters. A low distortion results in an excellent spurious free dynamic range (SFDR), which is required for communication systems such as mobile phone base stations. The succeeding stages use 5 bit and 4–6 bit in the last stage. The different track-and-hold stages run on different clocks to allow the optimum usage of the sample period.

For a nominal resolution of 14 bit the ENOB ranges from 11.3 to 11.8 bit. Sample rates between 150 Ms/s and 400 Ms/s are available with power consumptions ranging from 400 mW at 80 Ms/s to 2 W for 400 Ms/s. The SFDR ranges from 80 to 95 dB. Special low-swing digital output stages (LVDS) are applied to suppress digital noise. Some calibration of these parts is necessary to avoid alignment errors between the stages.

Figure 8.61 shows the lay-out of a 16-bit coarse-fine converter.

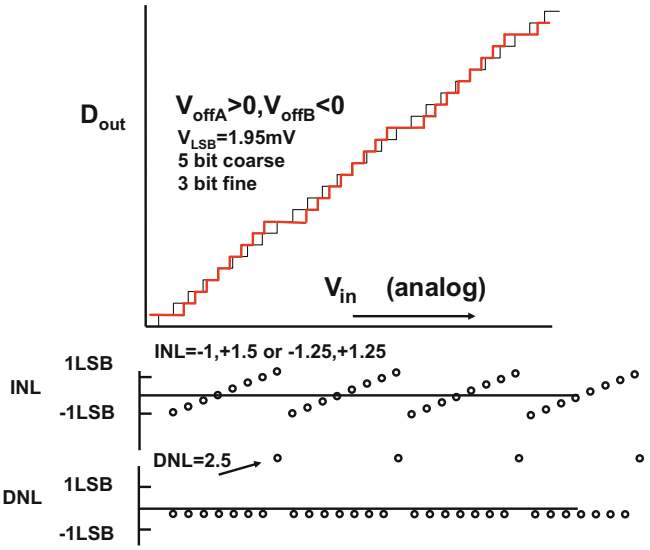


Fig. 8.62 Transfer curve, INL and DNL for a 3-bit fine section in a subrange analog-to-digital converter. With offsets in the buffers

Example 8.8. Figure 8.62 shows a part of the transfer curve in case the low-side buffer has a positive offset and the high-side buffer a negative offset. A plateau appears in the overall transfer curve with a DNL error corresponding to the offsets divided by V_{LSB}. If the offset appear the other way around, the transfer will look like Fig. 8.63. In this example two codes will be missing.

In an 8-bit 0.5-V subrange analog-to-digital converter the buffers between the 5-bit coarse and the 3-bit fine converter have 3 mV and −2 mV offsets. Sketch the resulting INL in case the buffers are switched directly or in a "monkey" way. What is in both cases the largest DNL?

Solution. Figure 8.64 shows the effect of the monkey-switching scheme. The maximum deviation of the INL has not changed, as it is determined by the offsets, but the local deviations at the ends of the fine range have largely disappeared. The remaining transfer curve consists of 8 codes that are together $|V_{off,A}| + |V_{off,B}|$ too large, followed by 8 codes that are the same amount too small. The DNL error per code is 5 mV/8 = 0.625 mV; related to V_{LSB} = 1.95 mV this would result in DNL = 0.3 LSB.

8.4 1-Bit Pipeline Analog-to-Digital Converters

Sub-ranging methods avoid the exponential hardware and power increase of flash conversion. Using more bits per subrange stage reduces the number of stages needed and the associated track-and-holds. If the first stage is implemented with

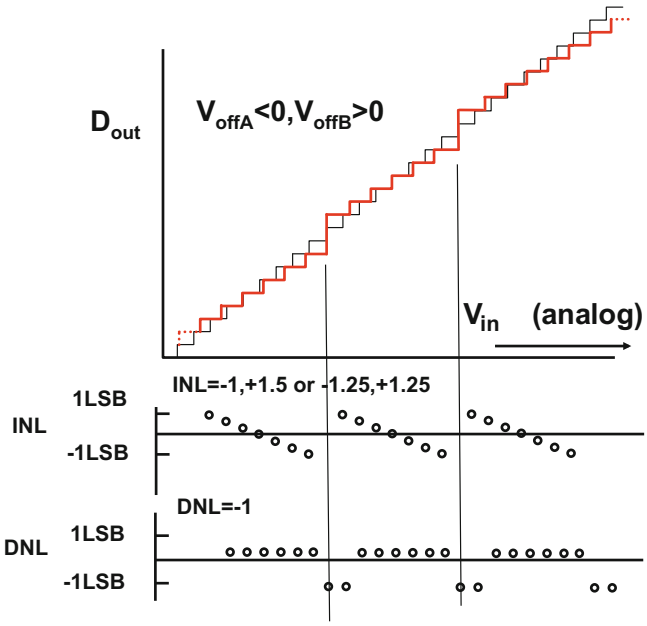


Fig. 8.63 Transfer curve, INL and DNL for a 3-bit fine section in a subrange analog-to-digital converter. With opposite offsets in the buffers

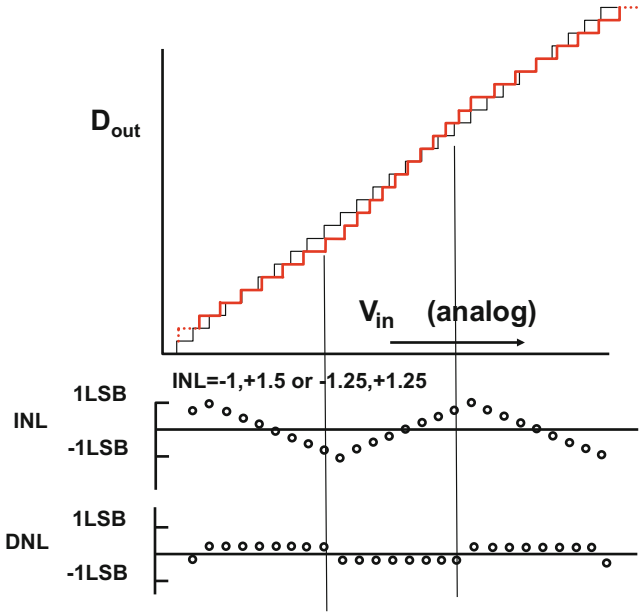


Fig. 8.64 Transfer curve, INL and DNL for a 3-bit fine section in a subrange analog-to-digital converter. Monkey switching reduces the DNL errors

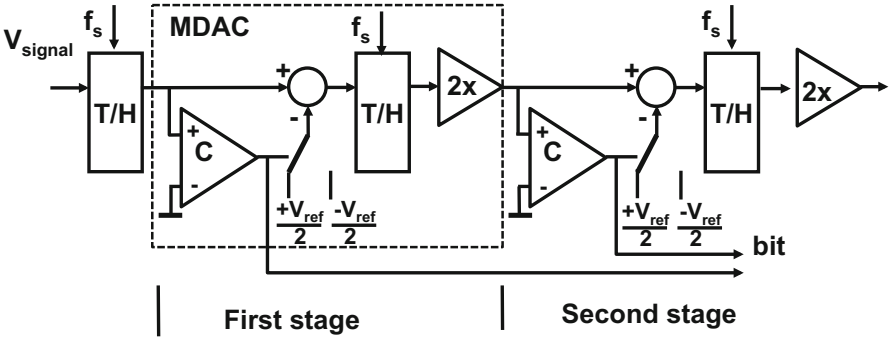


Fig. 8.65 Basic pipeline analog-to-digital converter

N_1 =3 to 5 bit, the intermediate gain can be of the order 2^{N_1}, thereby reducing the influence of errors in the succeeding stages. When a design is pushed to the limits of the technology, the maximum unity-gain bandwidth is determined by the available power and the capacitive load. More intermediate gain will then correspond to slower settling and lower sample rates, Fig. 8.52.

Reducing the subrange to the extreme results in subranges of one single bit. The direct benefit of this choice is that the digital-to-analog converter has a 1-bit resolution and is perfectly linear and will not contribute any distortion, see Fig. 7.52. Therefore a popular variant of the subrange analog-to-digital converter in CMOS technology is the 1-bit "pipeline" converter, Fig. 8.65 [210, 216]. This converter consists of a pipeline of N more-or-less identical 1-bit stages. For each bit of resolution there is one stage. Each stage (also called multiplying digital-to-analog converter, MDAC) comprises a T&H, a comparator connected to a one-bit digital-to-analog converter, a subtraction mechanism and a multiplication circuit.

With a conversion of one bit per stage, N stages are needed and N samples are simultaneously present in the converter. Additional time is needed for the digital reconstruction and optional error correction. Therefore the time between the first sampling of the signal and the moment the full digital value becomes available at the output is rather long ($N + 3$ clock periods). The resulting delay, called latency, will impair the system performance if this converter is part of a feedback loop.

Next to the advantage of fast settling due to a low interstage gain (2 or less) the 1-bit digital-to-analog converter allows a linearity improvement. With only two output values, this digital-to-analog converter is by definition perfectly linear.

8.4.1 Multiplying Digital-to-Analog Converter

The operation of the pipeline converter can be viewed from different angles. A pipeline converter can be seen as the extreme form of subranging with one bit per

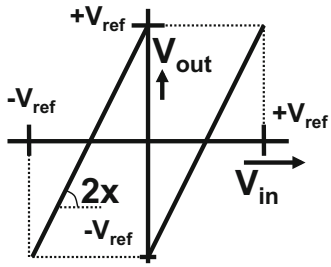


Fig. 8.66 Transfer characteristics for an MDAC in a pipeline analog-to-digital converter

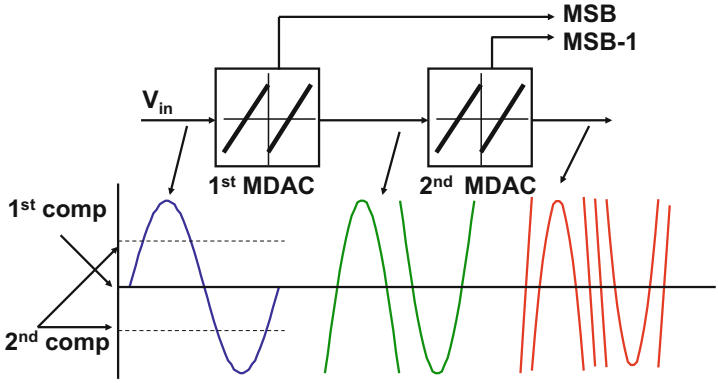


Fig. 8.67 The signal flow through the first two stages of a pipeline converter. Each MDAC produces a digital bit and an analog residue

subrange processed in an MDAC. From another point of view, the MDAC is essentially the kernel of the successive approximation algorithm from Fig. 8.90, except for the multiplication. The pipeline converter can therefore also be understood as a successive approximation converter, where every separate approximation step is implemented in dedicated hardware and the residue is amplified.

The transfer function of the input to the output of a single MDAC stage is shown in Fig. 8.66. The circuit multiplies the input signal by a factor 2. This would inevitably lead to an output voltage exceeding the voltage handling capability. A comparator determines whether the input is higher or lower than the middle of the reference voltages. The comparator decision leads to a subtraction or addition of the reference value. This results in a shifting down or up of the two halves of the multiply-by-2 curve. The result is passed to the next stage. For the i-th stage:

$$V_{out,i} = 2 \times V_{in,i} - D_i V_{ref} \tag{8.26}$$

where $D_i = -1, +1$ is the decision bit. Now the output signal uses the same range as the input signal. Every MDAC produces 1 bit and an analog residue, see Fig. 8.67. Simple cascading of N MDAC stages leads to an N-bit analog-to-digital converter.

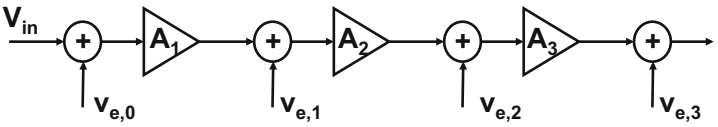


Fig. 8.68 Error sources in a chain of amplifiers add up. The input-referred error caused by an each error along the chain is divided by the total amplification from the input to the error

For a linear stage with an analog input and output signal, the transfer function is represented by a pair of straight lines.

When every stage reduces one bit, the remaining residue will have a maximum amplitude of half the value of the input signal. One-bit pipeline stages can therefore at maximum multiply by a factor of two.

The factor two gain implies that errors after the first MDAC will affect the input by half of their magnitude. For the errors behind the second MDAC, the errors are reduced by 4, etc. Deterministic errors create an input referred error in Fig. 8.68 of:

$$v_{in,det} = v_{e,0} + \frac{v_{e,1}}{A_1} + \frac{v_{e,2}}{A_1 A_2} + \frac{v_{e,3}}{A_1 A_2 A_3} + \ldots \tag{8.27}$$

Random errors, such as mismatch and noise(see Sect. 3.4), are normally independent errors,[11] they reduce in a root-mean-square sense. The random errors of the second stage and next stages add up to the error of the first stage as

$$v_{in,random} = \sqrt{v_{e,0}^2 + \frac{v_{e,1}^2}{A_1^2} + \frac{v_{e,2}^2}{A_1^2 A_2^2} + \frac{v_{e,3}^2}{A_1^2 A_2^2 A_3^2} + \ldots} \tag{8.28}$$

where $A_1, A_2, A_3, \ldots$ have a maximum value of 2.

The limited resolution-reduction per stage and the low amplification per stage make that the effect of errors of many stages results in a considerable effect at the input. This holds, e.g., for kT/C noise. Preferably a designer would like to reduce the size of the capacitor for the second-stage track-and-hold and to minimize the associated currents. However, the impact of the second stage noise does not allow such drastic reductions. In many designs a similar size capacitor is used for the first three stages. The total input referred kT/C noise is then $21/16\times = 1.3125\times$ the first stage noise. For the fourth and successive stages the requirements on the gain and accuracy also drop by a factor two per stage. Here "stage-scaling" allows the reduction of the capacitors and currents in these stages resulting in power-efficient designs [217].

[11] Be aware of bias noise and power supply variations, these can produce dependent errors.

8.4.2 *Error Sources in Pipeline Converters*

The choice to operate the sections at one-bit resolution has a number of consequences. The coarse analog-to-digital converter becomes a comparator. The digital-to-analog converter reduces in this scheme to a single bit converter deciding between a positive and a negative level. Two levels are connected with a straight line and are by definition perfectly linear. This solves the most important accuracy and distortion problem in the subrange digital-to-analog converter. In the specific 1-bit architecture of Fig. 8.65 there is no redundancy or over-range. And although the problem of an N-bit accurate digital-to-analog converter has been solved, the issue of connecting the coarse range to the fine range still exists for every MDAC and its neighbors. In this implementation, where the ranges of all digital-to-analog converters are fixed to $+V_{ref}, -V_{ref}$, the accuracy problem is in the exact $2\times$ amplifier and in the exact comparator decision.

The two important errors in this architecture are gain errors and comparator offset, Fig. 8.69. At the moment the transfer curve exceeds V_{ref} either on the positive side or on the negative side, the signal is out of the acceptable input range of the succeeding stage. If, on the other hand, the transfer curve does not reach the reference values, not all of the digital output codes will be used. Both types of errors result in loss of decision levels or missing codes in the transfer characteristic. Compare the same effect in coarse-fine conversion shown in Figs. 8.62 and 8.63.

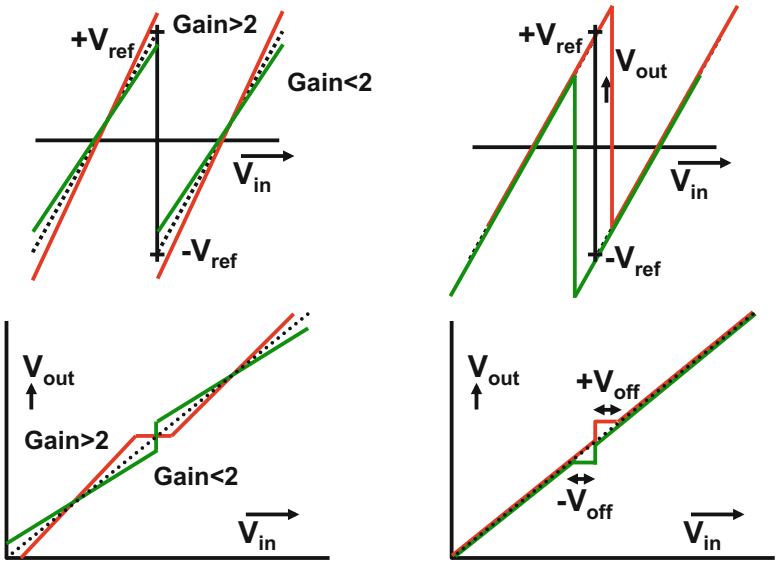


Fig. 8.69 Two errors in a pipeline stage. *Left*: the gain is not equal to 2 and *right*: the comparator suffers from a positive or negative offset. A converter with these errors in the first stage has an overall transfer characteristics as shown in the lower plots

The comparator offset and the gain errors must together remain under $0.5V_{LSB}$ to obtain a correct transfer curve.[12] With some offset compensation this condition can be reached for 10–12 bit accuracies. The comparator offset problem is more effectively solved in 1.5 bit pipeline converters, Sect. 8.5.

8.4.3 Multiply-by-Two Errors

The multiply-by-two operation in a pipeline converter is crucial to the overall accuracy. The following stages expect that the signal swings exactly between the positive and negative reference voltages. Any error creates similar problems as a misalignment between a coarse and a fine converter.

The process starts by sampling the signal on two equal capacitors C_1, C_2, see Fig. 8.70. In the multiply phase all charge is transferred to C_2 and depending on the comparator decision, V_{ref} is added or subtracted:

$$V_{out} = \frac{C_1 + C_2}{C_2} V_{in} \pm \frac{C_1}{C_2} V_{ref} \tag{8.29}$$

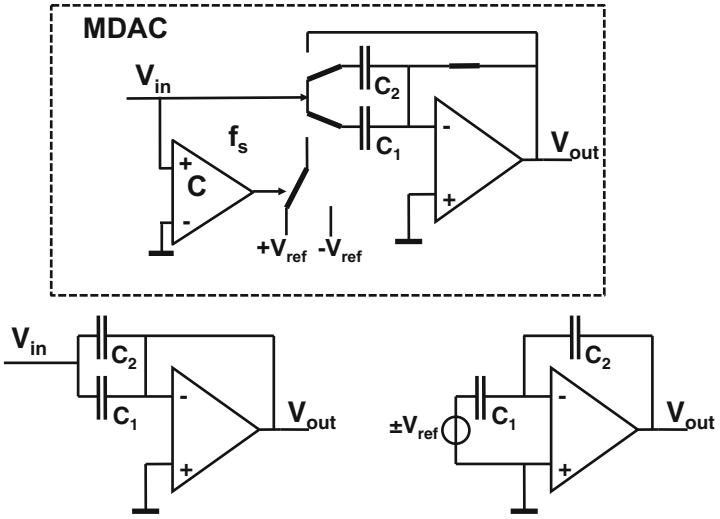


Fig. 8.70 The multiply-by-two operation in a pipeline converter is implemented by sampling the signal on two equal capacitors (*left*). In the multiply phase (*right*) all charge is transferred to C_2 [163]. Note that in this implementation the reference voltage is multiplied by 1 and therefore appears here as $\pm V_{ref}$

[12] This is not a sufficient condition as timing errors and the errors in the succeeding stages are ignored.

If $C_1 = C_2$, this operation results in an exact multiplication by two and a shift into the correct range, as desired in Fig. 8.66. A deviation in gain creates either missing range on the analog side or missing codes on the digital side. In this implementation, the signal sample voltage is multiplied by 2, while the transfer of the reference voltage is only 1×. This is a difference in reference voltage handling between the generic scheme of Fig. 8.65 and the implementation in Fig. 8.70.

As discussed in Sect. 3.4.4 the **offset of the opamp** is corrected if during the initial sampling the virtual ground is generated by the opamp itself.

Errors in the capacitor ratio can be the result of technological problems or deviations caused by the lay-out environment, see Sect. 5.2.2. The overall DC-accuracy requirement is translated in a minimum mismatch error for the capacitor ratio of:

$$\epsilon_{cap} = \frac{C_1 - C_2}{C_1 + C_2} \ll 2^{-N+i} \tag{8.30}$$

where $i = 1, 2, \ldots N$ is the stage number.

Minimizing the capacitor error requires the lay-out of an accurate structure, see, e.g., Fig. 7.46. Unfortunately the random errors can only be reduced via larger area: Eq. 5.32. Example 8.9 provides an insight in the required size. Several examples show that with careful lay-out an accuracy of 14-bit is possible with calibration [218, 219].

Next to minimizing the capacitor mismatch in a technological manner, several schemes try to mitigate this error in an algorithmic way. Li et al. [220] proposes the ratio-independent technique. A signal is sampled in a capacitor and in a second cycle stored in an intermediate capacitor. In a third cycle the first capacitor takes another sample, and in a fourth cycle the first sample is retrieved and added to the second sample.

Song et al. [42] and Chiu et al. [221] present a scheme where the multiply-by-two is executed in two phases: in the first phase the sample and amplify cycle is performed with the capacitors arranged as shown in Fig. 8.71 (upper). The difference comes with the second phase where C_1 and C_2 are interchanged, Fig. 8.71 (lower). The results of both phases are averaged yielding a reduced effect of the capacitor mismatch. Unfortunately these schemes cost hardware and several clock cycles.

Another approach to calibrate the errors in a fast pipeline converter is in combination with a slow high-precision analog-to-digital converter. Synchronization of both converters is achieved by a track-and-hold circuit or post processing [222].

An opamp gain error is the result of insufficient opamp gain, capacitor mismatch, or switch charge injection, see Sects. 3.3.2 and 5.2.2. Each of these errors affects the transfer of the residue signal and must remain within a fraction ($\alpha = 0.1 - 0.2$) of an LSB. The opamp DC-error after settling is given by feedback theory and must be smaller than the overall required accuracy:

$$V_{out}\left(1 + \frac{C_1 + C_2}{A_0 C_2}\right) = \frac{C_1 + C_2}{C_2} V_{in} \pm \frac{C_1}{C_2} V_{ref} \tag{8.31}$$

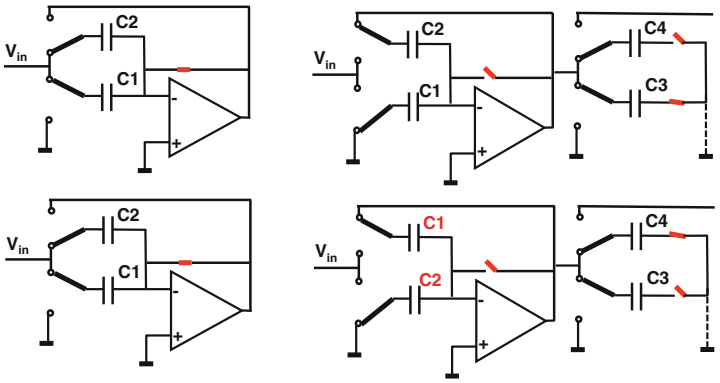


Fig. 8.71 The multiply-by-two operation is executed twice: and the results are averaged [42, 221]

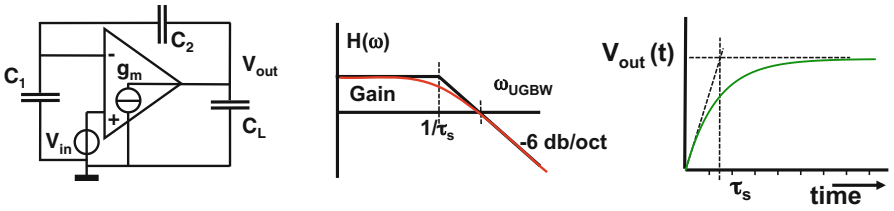


Fig. 8.72 *Left*: analysis diagram, *Middle*: frequency response, *right*: time response on a step input

$$V_{out} = \frac{(C_1 + C_2)V_{in} \pm C_1 V_{ref}}{C_2 + (C_1 + C_2)/A_0}$$

$$\epsilon_{DCgain} = \frac{(C_1 + C_2)}{A_0 C_2} \ll \alpha 2^{-N+i} \quad \rightarrow \quad A_0 > 2^{N-i+1}/\alpha$$

with resolution N and the DC-gain is A_0. Typical candidates for an opamp topology are: telescopic (single pole, single current, but limited output range), folded-cascode (multiple current paths, parasitic pole in cascode), and two-stage Miller (high gain, multiple currents, but dominant Miller pole), where each of these topologies can be gain-boosted [223].

For stages further on in the pipeline the demands are less critical. The stage number $i = 1, 2, \ldots N$ illustrates the option for stage-scaling. The amplification of the first stage is the highest. The amplification requirement holds for the entire output range, so any saturation effects in the output range will cause loss of accuracy. Alternative ways to reduce the requirements on the gain are based on reduction of the output swing as in Fig. 3.26 [43] and multiple gain steps [224].

The limited settling speed of the operational amplifiers determines the speed of the pipeline converter. **Limited settling** causes errors comparable to static gain errors. Figure 8.72 shows the analysis diagram of the settling behavior in the amplification mode. Straight-forward analysis gives

$$\frac{v_{out}}{v_{in}} = \frac{g_m(C_1 + C_2)}{j\omega(C_1C_2 + C_1C_L + C_2C_L) + g_mC_2} \tag{8.32}$$

The metrics of the asymptotes are derived as

$$\text{Gain} = \frac{1}{\beta} = \frac{C_1 + C_2}{C_2} \quad \omega_{\textbf{closed-loop}} = \beta\omega_{UGBW} \quad \tau_s = 1/\omega_{\textbf{closed-loop}} \tag{8.33}$$

where $\beta = C_2/(C_1 + C_2)$ is the feedback factor. The time constant for the settling of the operational amplifier is $\tau_s =$ and equals the inverse of the closed-loop bandwidth $\omega_{\textbf{closed-loop}}$.

In case of a single-pole transconductance amplifier, such as a telescopic amplifier:

$$\omega_{\text{UGBW}} = \frac{g_m}{(C_L + C_1C_2/(C_1 + C_2))} \quad \tau_s = \frac{(C_1C_2 + C_1C_L + C_2C_L)}{g_mC_2} \tag{8.34}$$

The unity gain bandwidth equals the transconductance divided by the capacitive load: C_L parallel to the series connection of C_1 and C_2. The implicit assumption is that this amplifier is single pole, in case of a two-stage Miller compensated amplifier, the UGBW equals: $\omega_{\text{UGBW}} = g_m/C_{Miller}$.

The time constant for the settling of the operational amplifier τ_s and the available settling time T_{settle}, which is preferably a large fraction of T_s, define the settling error:

$$\epsilon_{settle} = e^{-T_{settle}/\tau_s} \ll 2^{-N+i-1} \tag{8.35}$$

$$f_{UGBW} > \frac{(N - i + 1)\text{Gain}\ln(2)}{2\pi T_{settle}} = \frac{f_s(N - i + 1)\text{Gain}\ln(2)}{2\pi(\text{duty cycle})} \tag{8.36}$$

where the duty cycle equals T_{settle}/T_s.

If the converter is operated in a simple mode where the comparator is active in one clock phase and the settling takes place in the other clock phase, T_{settle} must be chosen according to the minimum duty cycle of the clock, pushing up the required UGBW of the opamp. And increasing the UGBW of the operational amplifier requires a lot of power. If a more complex timing sequence is used, the settling time for the operational amplifier and the time needed for the comparator can be better controlled allowing more time for the settling and saving of some opamp power.

The above-mentioned error sources must not exceed a total of $0.5V_{LSB}$. The balance between the error sources depends on the required specifications and available technology: high accuracy requires high gain and low mismatch. High speed requires simple and fast transconductance stages.

Example 8.9. An MDAC is used for an exact multiply-by-2 converter, Fig. 8.70 with a transconductance as an amplifier. Suppose the switches and opamp are ideal, calculate the value for C_1 and C_2 for 10-bit performance and 2-sigma yield?

Solution. The required gain function for proper operation of this MDAC is: $A_v = (C_1 + C_2)/C_2$. The expectation values for the capacitors are $E(C_1) = C, E(C_2) = C$ and the variances for both are: $\sigma_{C1}^2, \sigma_{C2}^2$. The variance of the difference $\Delta C = C_1 - C_2$ is found as $\sigma_{\Delta C}^2 = \sigma_{C1}^2 + \sigma_{C2}^2$.

The variance in the gain can be found by applying Eq. 5.11 to the nominal gain:

$$\sigma_{Av}^2 = \left(\frac{\partial A_v}{\partial C_1}\right)^2 \sigma_{C1}^2 + \left(\frac{\partial A_v}{\partial C_2}\right)^2 \sigma_{C2}^2 = \left(\frac{1}{C_2}\right)^2 \sigma_{C1}^2 + \left(\frac{-C_1}{C_2^2}\right)^2 \sigma_{C2}^2$$

For a multiply-by-2 MDAC, the capacitors will be nominally equal and show identical statistical properties. In that case:

$$\sigma_{Av}^2 = \frac{\sigma_{C2}^2}{C^2}$$

If the yield criterium is 0.5 LSB, then $\sigma_{C2}/C_2 = \sigma_{Av} < 0.25LSB = 2^{-N-2} = 2^{-12}$. Table 5.6 provides the relation between the relative capacitor error and the capacitor value. With $\frac{\sigma_{\Delta C}}{C} = \frac{A_C}{\sqrt{\text{C in fF}}}$ and A_C =0.5%$\sqrt{fF}$ the capacitor value is found as: $C_1 = C_2 > 0.4$ pF.

Example 8.10. An MDAC is used for an exact multiply-by-2 converter, see Fig. 8.70. Suppose the switches are ideal, calculate an expression for the input referred noise.

Solution. During sampling kT/C noise is stored on the two capacitors C_1 and C_2: $v_{in,sample-noise}^2 = kT/(C_1 + C_2)$. This voltage corresponds to a charge: $q_{noise}^2 = kT(C_1 + C_2)$, which is transferred to C_2 and creates an output noise $v_{out,sample-noise}^2 = kT(C_1 + C_2)/C_2^2$.

However, during the amplification phase the opamp generates noise. The transfer function of the transconductance amplifier is

$$\frac{v_{out}}{v_{in}} = \frac{(C_1 + C_2)}{C_2} \frac{1}{1 + j2\pi f C_1/g_m}$$

The input transconductance g_m generates noise: $v_{gm,noise}^2 = 8kTBW/3g_m$. Forming the power integral gives as a result: $v_{out,track-noise}^2 = (2/3)kT(C_1 + C_2)/(C_1C_2)$ or $v_{out,track-noise}^2 = kT(C_1 + C_2)/(C_1C_2)$, where the noise of the transconductance is modeled as a full resistive noise $4kTBW/g_m$. More informally: the noise on the output equals the kT/C contribution of the capacitive load on that node. The load is the series connection of the two capacitors C_1 and C_2.

The total output referred noise is now $v^2_{out,total-noise} = kT(C_1 + C_2)^2/(C_1C_2^2)$ and to the input referred

$$v^2_{in,total-noise} = \frac{v^2_{out,total-noise}}{Gain^2} = \frac{kT(C_1 + C_2)^2/(C_1C_2^2)}{(C_1 + C_2)^2/C_2^2} = \frac{kT}{C_1}$$

After setting $C_1 = C_2 = C$:

The sampled noise on the parallel input capacitors is: $v^2_{in,sample-noise} = kT/2C$, after amplification by-2 or in power by-4: the output noise is $v^2_{out,sample-noise} = 2kT/C$.

During tracking there is noise on the output on the series connection of C_1 and C_2 or on $C/2$: $v^2_{out,track-noise} = 2kT/C$. Referred back to the input means dividing by the power-gain: $v^2_{in,track-noise} = kT/2C$. The total input referred noise after the track, amplify, and hold operation is

$$v^2_{in,total-noise} = v^2_{in,sample-noise} + v^2_{in,track-noise} = \frac{kT}{2C} + \frac{kT}{2C} = \frac{kT}{C}$$

This noise power in Fig. 8.70 consists of two fundamentally different contributions. On one hand, $v^2_{in,sample-noise}$ is a sampled noise power in a bandwidth $0, \ldots, f_s/2$, while $v^2_{out,track-noise}$ is a time-continuous noise running up to f_{UGBW}, which (theoretically) can be a completely different frequency range. Therefore the noise density in $0, \ldots, f_s/2$ is composed of the sampling contribution kT/Cf_s, and the time-continuous contribution $kT/2Cf_{UGBW}$.

Further analysis takes the lower noise in transconductances into account.

Compare this result to [226] and note that the switching sequence in [226] differs.

8.4.4 Reduced Radix Converters with Digital Calibration

In order to circumvent the accuracy problems in Fig. 8.69, Karanicolas et al. [227] proposes to use a gain factor which is deliberately set smaller than 2. The gain factor is chosen to avoid range overflow either by comparator offset or gain errors, see Fig. 8.73. With less than 1-bit resolution per stage, more stages M are needed than the resolution N. This topology is in fact a conversion with a radix or base less than 2.

The actual gain is not precisely known. In order to reconstruct the input from the (digital) output value the following transfer equations are obtained for the left and right part of the transfer curve in Fig. 8.73 (left):

$$V_{in} < 0, \rightarrow V_{out} = \frac{V_A - V_B}{V_{ref}} V_{in} + V_A$$

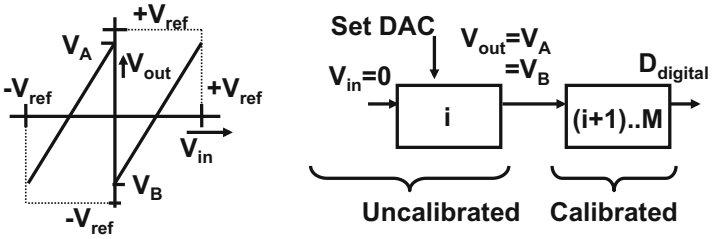


Fig. 8.73 The gain in the stage is deliberately made smaller than 2. The converter is calibrated starting at its back end and working towards the front-end [227]

$$V_{in} > 0, \rightarrow V_{out} = \frac{V_A - V_B}{V_{ref}} V_{in} + V_B \tag{8.37}$$

This reconstruction requires that the intersection points with the vertical axis V_A and V_B are known. If these (digital) values are known the above equations allow to "glue" the left and right part of the curve together to form a straight transfer curve. These values are measured with the last part of the converter consisting of sections $i+1, \ldots, M$ that have been calibrated before. During the digital calibration cycle the values of V_A and V_B are measured as digital codes with the help of the calibrated part of the converter. In calibration mode the input of section i is grounded $V_{in} = 0$ and the comparator is set to a negative result, Eq. 8.37 (upper). The output of section i will be equal to the value V_A. Similarly V_B is obtained with the comparator set positive. Now that the digital values of V_A and V_B are known, the transfer curve of section i can be obtained and used to measure the voltages V_A and V_B of section $i-1$. The calibration process starts with the last section and continues until the first section is reached. The values of V_A and V_B must be stored and kept in non-volatile storage.

Example 8.11. Discuss the calibration of a reduced radix converter based on Fig. 8.74, determine the values V_A and V_B.

Solution. Figure 8.74 gives an example of the calibration. Before the pipeline converter is ready for use, the calibration cycle has to determine the digital values for V_A and V_B. In this example the last converter stage is a 4-bit flash converter and V_A and V_B are measured as digital values of 14 and 2 by switching the input signal to $V_{in} = 0$ and toggling the comparator output. In fact these two codes correspond to one and the same input voltage (in this example $V_{in} = 0$) for two settings of the digital-to-analog converter/comparator. Therefore the conversion curves corresponding to $D_i = 0$ and $D_i = 1$ can be aligned on these points. For $D_i = 0$ code 14 corresponds to $D_i = 1$ code 2. After calibration a negative input signal will be measured by codes $D_i = 0$ code: 2,...,14. A positive signal continues with $D_i = 1$ code 2,...,14. The resulting converter has 26 trip levels and is close to a 5-bit range.

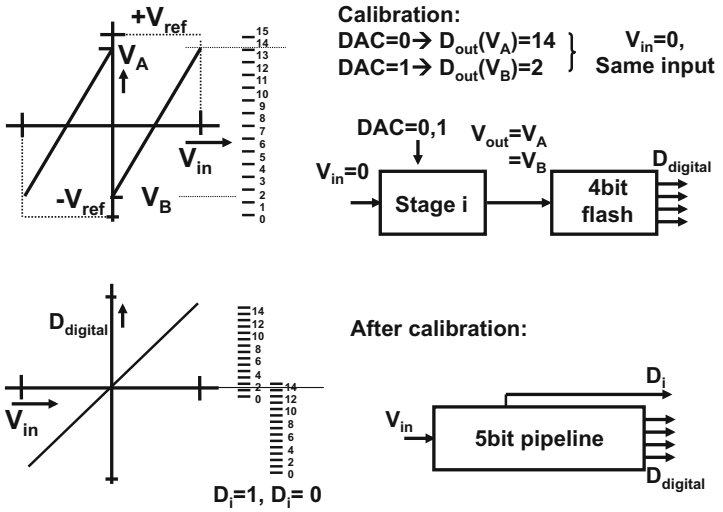


Fig. 8.74 Example of pipeline calibration

The local reference and the quality of the first subtraction or addition must still achieve the accuracy level of the converter. The reconstruction of signal samples is now a simple addition of the digital codes weighted with the decision of the comparator in each section. This calibration relaxes the accuracy requirements on the overall gain. Also comparator offset is not critical, as long as V_A and V_B stay within the reference range.

8.5 1.5 Bit Pipeline Analog-to-Digital Converter

The straight-forward pipeline converter of Fig. 8.65 requires full and accurate settling in all stages. Digital calibration is needed when the gain factor is reduced. Still the comparator offset is not corrected in the scheme of Fig. 8.73.

The 1.5 bit pipeline converter allows a more extensive error correction as in each stage two comparators are used, see Fig. 8.75. The trip levels of these comparators are typically located at 3/8 and 5/8 of a single sided reference or at $\pm 1/4$ of a differential reference, see Fig. 8.76 [38, 221, 225, 228–230]. These decision levels split the range in three more or less equivalent pieces, thereby relaxing the constraints on the input range of the circuit. In case of a double positive decision of both comparators, the digital-to-analog converter will set the switches to subtract a reference value. In case of two negative decisions the digital-to-analog converter will add a reference value and in case of a positive and a negative decision, the signal will be passed on directly to the amplifier. Every stage therefore is said to generate 1.5 bit. The output voltage of an MDAC ideally is

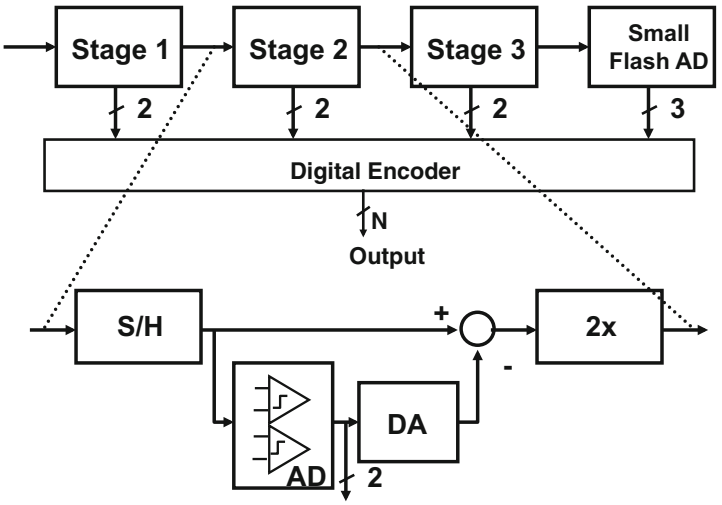


Fig. 8.75 The architecture of a 1.5-bit pipeline analog-to-digital converter is derived from the coarse-fine architecture

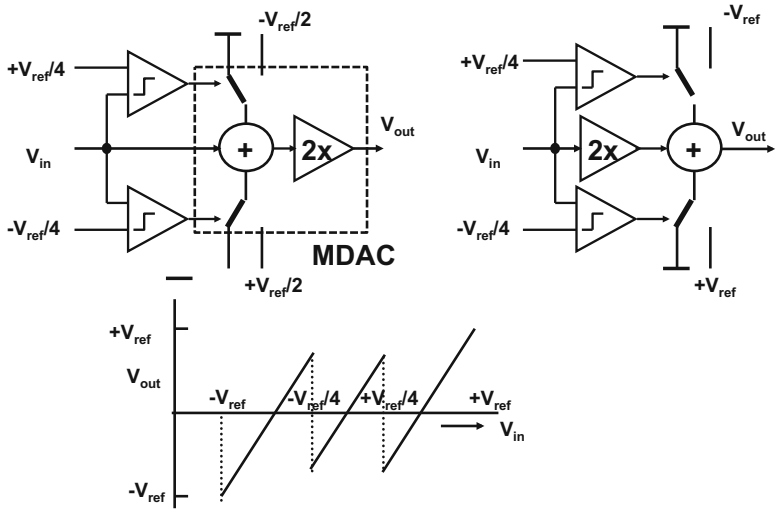


Fig. 8.76 Basic operation of a 1.5-bit pipeline analog-to-digital converter. The section in the *dotted box* is called a multiplying digital-to-analog converter (MDAC)

$$V_{out} = 2V_{in} + \begin{pmatrix} +V_{ref} \\ 0 \\ -V_{ref} \end{pmatrix} \tag{8.38}$$

where both V_{in}, V_{out} stay within $-V_{ref}, .., +V_{ref}$.

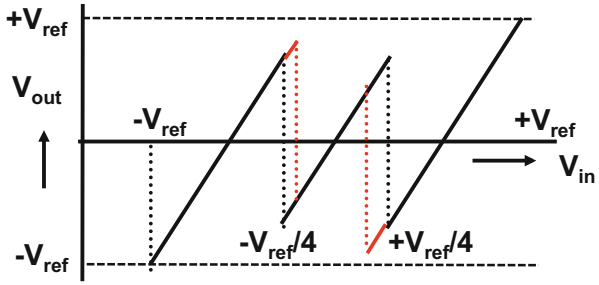


Fig. 8.77 The two comparators have offsets. As long as the curve remains between the maximum outputs $+V_{ref}$ and $-V_{ref}$ the decision can be corrected

The digital-to-analog converter generates three levels. In a fully differential design these levels can be made without loss of accuracy: zero, plus, and minus the reference voltage. The last two levels are generated by straight-forward passing the reference voltage or twisting the connections.

The typical transfer curve in Fig. 8.76 shows that negative signals are shifted upwards and positive signals are shifted downwards in the plot with respect to the values around zero.

If the comparators show offset, the switching points in the output plot of Fig. 8.77 shift. As long as the extreme values of the transfer curve remain inside the maximum values of the output, there is no information lost and an option exists to correct the decision.

8.5.1 Redundancy

Figure 8.78 shows the redundancy mechanism of a 1.5 bit scheme. The upper curve shows the voltage range of seven stages and ideal comparator levels. An input voltage equal to $V_{in} = 0.6V_{ref}$ clearly exceeds the upper comparator threshold at $0.25V_{ref}$ in the first stage. Therefore the first coefficient is determined to be $a_{N-1} = 1$. The residue is formed as $V_{out,1} = 2 \times V_{in} - V_{ref} = 0.2V_{ref}$. This value generates in the second stage a $a_{N-1} = 0$ decision. For this stage: $V_{out,2} = 2 \times V_{in} = 0.4V_{ref}$ In the following stages the rest of the conversion takes place. The result "77" equals $77/128 = 0.602V_{ref}$.

In the lower trace of Fig. 8.78 the comparator level of the third section is moved to $0.45V_{ref}$. It is easy to see how the redundancy scheme corrects for this error. The redundancy eliminates the need for accurate comparators, but leaves full accuracy requirements on the switched-capacitor processing stages.

See also the remarks for the RSD converter in Fig. 8.117.

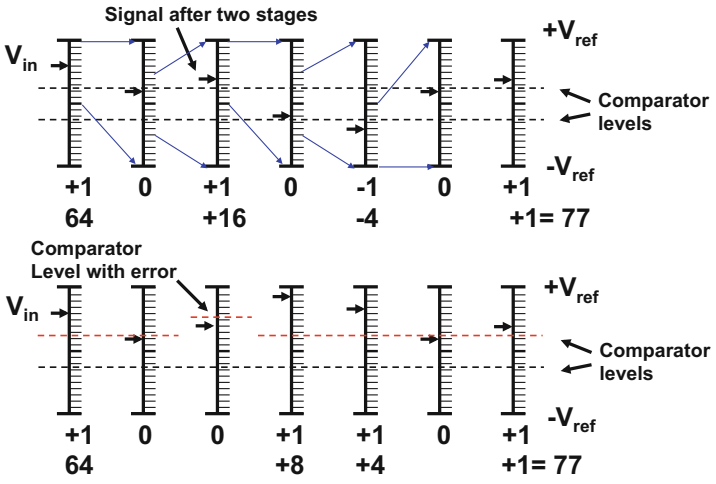


Fig. 8.78 1.5-bit pipeline analog-to-digital converter uses the redundancy of two decisions per stage. The *upper sequence* shows the ideal behavior. In the *lower sequence* one comparator level is shifted, e.g. due to offset. The redundancy corrects the mistaken decision

8.5.2 Design of an MDAC Stage

Like in most CMOS analog-to-digital principles, also 1.5-bit pipeline converters need good capacitors, high-performance operational amplifiers, and low-leakage CMOS switches to implement high-quality track-and-hold stages and to create an N-bit pipeline analog-to-digital converter.

The first choice that has to be made is the signal swing. Operating the converter in differential mode is preferred as it doubles the amplitude with respect to a single-sided approach. Moreover it minimizes even order distortion, power supply, and substrate noise influence. However, analog-to-digital converters that take their input directly from external sources will normally see a single-sided signal. Also many RF and filter circuits before a converter are designed single-sided to allow optimum and/or minimum use of components. In both cases it is desirable to have an as large as possible input signal swing.

A popular implementation is shown in Fig. 8.79 (left). This Multiplying DAC (MDAC) configuration allows a large signal swing without any repercussion for the input of the operational amplifier. Its inputs remain at virtual ground level thereby avoiding common mode problems (such as in telescopic amplifiers).

The MDAC is largely identical to the MDAC in Sect. 8.4.1, except for the additional comparator and the additional reference voltage.

During the track phase in Fig. 8.79 (left), the switch S_3 puts the operational amplifier in unity gain feedback (dotted switch positions). The capacitors C_1 and C_2 are connected in parallel to the input signal. When the transition to the hold-amplify-phase occurs (bold switch positions), switch S_1 creates via C_2 the feedback

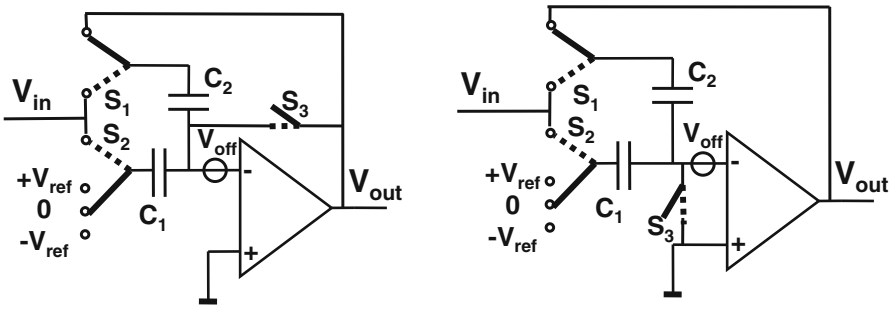


Fig. 8.79 A track-and-hold with multiply-and-subtract stage (MDAC) used in 1.5-bit pipeline converters. Two alternative schemes are shown for creating a ground node in pipeline converters. For the track phase the switches are in the dotted positions. In the hold-and-amplify phase the switches in the drawn positions

path for the amplifier. S_2 switches capacitor C_1 to one of three reference voltages, depending on the result of the two comparators. The charge corresponding to the input signal on C_1 is transferred to C_2. In addition a charge packet corresponding to the chosen reference voltage is moved into C_1 and C_2. The overall transfer of this circuit is

$$V_{out} = \frac{(C_1 + C_2)}{C_2} V_{in} + D \times \frac{C_1}{C_2} V_{ref} \tag{8.39}$$

where D represents the decision from both comparators and can take the values $+1, 0, -1$. In fact this is a form of ternary logic.

When $C_1 = C_2$ the signal is multiplied by two, while the additional reference is added, ignored, or subtracted. This principle has a lot of similarity with algorithmic analog-to-digital converters, see Sect. 8.7.

In Fig. 8.79 (left) switch S_3 sets the opamp in unity-gain mode in order to create a virtual ground during track mode. The opamp and the switches operate as described for auto-zeroing comparators in Sect. 8.1.9 with an increased noise level due to the sampling of the high-frequency parts of the input referred noise of the opamp. The switch S_3 can be moved to a position where it shorts the input terminals of the amplifier directly to ground, Fig. 8.79 (right) [231]. This real ground node for the sampling process removes the input-referred opamp noise. Now the input referred offset of the opamp V_{off} is introduced into the amplification phase, leading to

$$V_{out} = \frac{(C_1 + C_2)}{C_2} (V_{in} + V_{off}) + D \times \frac{C_1}{C_2} V_{ref} \tag{8.40}$$

The curve in Fig. 8.77 will be shifted vertically by the amplified input referred offset. This will lead to loss of range on one side of the transfer curve. These configurations trade off the offset cancellation of the scheme in Fig. 8.79 (left) for lower noise in Fig. 8.79 (right).

Figure 8.79 (right) has two additional advantages: the input sampling does not need an opamp to keep the virtual ground steady at high input frequencies. Direct connection to ground allows wide tracking bandwidths for systems requiring down-sampling, see Sect. 2.3.2. The second advantage is that the opamp can be used to perform the amplification function of another section: "opamp-sharing."

Many performance criteria for designing a 1.5-bit MDAC match those for the 1-bit MDAC of Sect. 8.4.3. The amplification stage has to fulfill similar demands as in the 1-bit pipeline converter. The topology of the opamp depends on the specifications. Low accuracy and high speed will push towards simple transconductance stages, while high accuracy may involve (folded) cascode amplifiers. The desired DC-gain is set by the overall resolution in Eq. 8.32, and the settling in Eq. 8.36.

The required capacitor accuracy is given by Eq. 8.30. Plate or fringe capacitors were described by Eq. 5.32, leading to:

$$\frac{\sigma_{\Delta C}}{C} = \frac{A_C}{\sqrt{\text{C in fF}}} < V_{pp} 2^{-N}$$

With $A_C = 0.5\%\sqrt{fF}$, a 2.5 pF capacitor will result in $\sigma_{\Delta C}/C = 10^{-4}$. Normally these values allow to reach 12–14 bit capacitor matching. Singer and Brooks [218] reports a 14-b, 10-Ms/s calibration-free pipeline converter based on 4 pF capacitors, with DNL = 0.7 LSB and a THD = −87 dB performance. Ali et al. [219] also claims 14-bit capacitor-matching performance without the need for calibration for a total input capacitor of 6 pF.

Moreover, high-resolution converters require the track-and-hold capacitor to fulfill the kT/C noise limit. The input referred noise of the opamp must be added and during the switching kT/C contributions of the following section must be included. Depending on the required specifications and the design of the opamp, a commonly used design guideline is to use an excess factor of ≈ 3 below the quantization noise, losing 1.3 dB of SINAD:

$$\frac{3 \times kT}{C_1} < \frac{V_{LSB}^2}{12} = \frac{V_{pp}^2 2^{-2N}}{12} \tag{8.41}$$

where V_{pp} represents the peak–peak value of the input signal.

The capacitor choice is a determining factor in the design. It determines the quality of the (uncalibrated) gain. The total noise accumulates as in Eq. 8.28 and Fig. 3.24. However, the capacitor value also determines the slew current setting of the opamp and thereby most of the analog-to-digital converters power. The bias current I_{bias} must be able to generate sufficient transconductance to reach the UGBW of Eq. 8.36. Moreover the slew rate requirements must be met:

$$I_{bias} > I_{slew} = C_{load} \frac{dV}{dt} \tag{8.42}$$

Table 8.4 Comparison of the main parameters for an MDAC with 8 and 12 bit resolution in a 1.5-bit pipeline stage (inspired by ISSCC workshops)

Parameter	Equation	N=8 bit	N=12 bit
Sample rate	f_s	100 Ms/s	20 Ms/s
LSB size	$V_{LSB} = 2^{-N} V_{pp}$	3.9 mV	0.24 mV
Capacitor based on noise	8.41: $3 \times kT/C_1 < V_{LSB}^2/12$	9.5 fF	2.4 pF
Capacitor matching	8.30: $\frac{\sigma_{\Delta C}}{C} \ll 0.25 \times 2^{-N}$	10^{-3}	6×10^{-5}
Capacitor based on matching and $A_C = 0.5\% \sqrt{fF}$	5.32: $\frac{\sigma_{\Delta C}}{C} = \frac{A_C}{\sqrt{\text{C in fF}}}$	25 fF	7 pF
Opamp gain	8.32: $A_0 > 2^N$	300 (50 dB)	5000 (74 dB)
UGBW ($T_{settle}/T_s = 0.5$)	8.36: $f_{UGBW} > \frac{f_s (N - i + 1) 2 \ln(2)}{2\pi 0.5}$	350 MHz	100 MHz
Slew current	8.42: $I_{slew} > C_{load}(dV/dt)$	10 µA	1100 µA

where the load capacitance C_{load} includes the feedback and parasitic capacitances as well as the input capacitance of the next stage. A safe measure for the voltage slope is $V_{max-swing}/\tau_{UGBW}$. The first stages are contributing most to the noise and inaccuracy budget, the following stages can be designed with relaxed specifications: stage-scaling.

Example 8.12. Determine the parameters for the first section of a pipeline stage for 8-bit resolution at 100 Ms/s and 12-bit resolution at 20 Ms/s, with a 1 V input range.

Solution. Table 8.4 lists the main parameters. Of course fine-tuning with a simulation is needed to incorporate the parasitic effects.

The total performance depends on noise, matching, and distortion contributions and must be balanced against power. The optimum choice depends on the application and the choices in this table are therefore only starting values for an application discussion.

8.5.3 Multi-Bit MDAC Stage

Judging from the number of published papers, for many years the pipeline converter is the most popular analog-to-digital converter.[13] Most effort is spent on converters in the range of 10–14 bit of resolution and 50–500 Ms/s sampling rates. These high-performance converters aim at the same markets as high-end subranging converters. Similar to the high-performance subranging converters the main problems are found in the design of the first track-and-hold circuit: constant switch resistance, the settling behavior of the opamp, and parasitics.

[13] Watch out for the successive approximation converter!

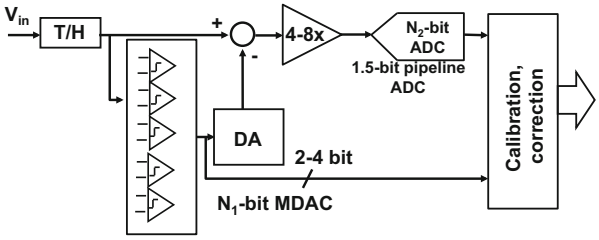


Fig. 8.80 Advanced pipeline concepts use a multi-bit MDAC as input stage

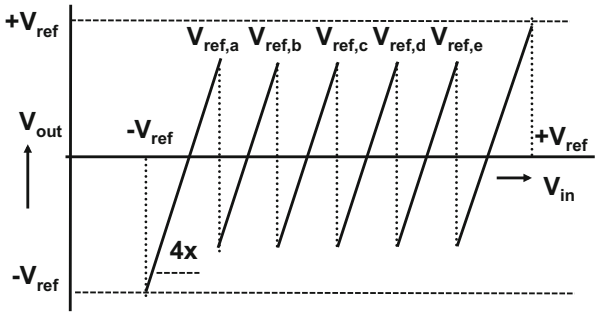


Fig. 8.81 Transfer curve for an MDAC stage employing five comparators and with a gain of 4×

With the first MDAC stages resolving 1.5-bit, the contribution of the second and third stages to the input referred noise is still significant. Often the same capacitor value for these stages is used and scaling is applied after the third stage. For those lower-resolution MDAC stages, the reduced capacitive load means lower currents and power saving is an obvious advantage, e.g. [217].

The alternative to 1.5-bit MDAC stages for the first conversion stage is to use more levels in the first stage and thereby directly allow stage-scaling in the lower-resolution MDACs. Basically the first stage is now the coarse converter in a sub-ranging architecture, while the fine converter is implemented as a 1.5-bit pipeline converter. See Fig. 8.80.

Most of these converters use input stages with 2.5–4 bit resolution, e.g. 6 levels [231] or 7 levels [39], and many other variants have been published [222, 232–236]. The analysis in [221, Appendix] indicates that 2.5 to 4 bit is an optimum range.

Figure 8.81 shows the transfer curve of a 4× amplifying input stage. The range is subdivided in six segments by five comparators. Again, the comparator offsets are correctable as there is margin for the curve before it passes $\pm V_{ref}$, where the next stage cannot handle the signal. The four-times amplification increases the time constant of the amplifier and a lower sampling speed will result unless the power is increased. Now also five digital-to-analog conversion levels must be created at full conversion accuracy or the INL/DNL will increase and distortion performance will drop.

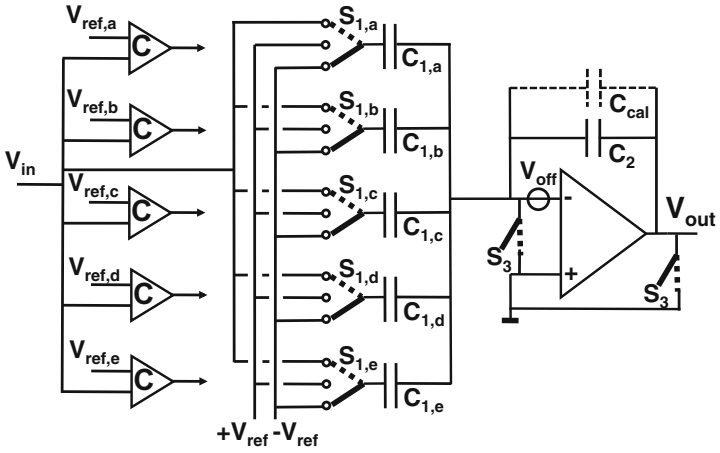


Fig. 8.82 Schematic diagram of a 5-level MDAC stage with 4× amplification, after [222, 236]. The circuit is shown in a simplified single-ended format, in a real design it is obvious that a differential topology must be used. The calibration capacitor is proposed in [236]

In Fig. 8.82 the digital-to-analog converter is implemented with a unary capacitor architecture. This allows an optimum performance as the unary architecture is less prone to DNL errors. Moreover, capacitors are in most CMOS technologies the best matching components. In this architecture the sampling capacitors are not used to perform a flip-around amplification. So the gain is given by $(C_{1,a} + \ldots + C_{1,e})/(C_2 + C_{cal})$. Again the accuracy of the multiplication is essential for achieving high resolution. In [236] a calibration capacitor is used to achieve an accurate matching between the input MDAC and the remaining stages. In a differential topology, both $+V_{out}$ and $-V_{out}$ are available thereby allowing also to use C_{cal} to subtract capacitance by connecting it to the $-V_{out}$ output. Tseng et al. [236] achieves 8.2 ENOB near $f_s/2$ input frequencies with a sampling rate of 320 Ms/s in 90 nm CMOS for 42 mW. In [222] a similar structure is used with flip-around and binary weighted digital-to-analog conversion. This design in a 0.35 µm technology reaches 11.4 ENOB at 20 Ms/s and 231 mW.

Figures 8.83 and 8.84 show the architecture and the lay-out of a 10-bit pipeline converter. The differential input is fed into a 2.5-bit MDAC with 4× gain, followed by 7 normal 1.5-bit stages and a 2-bit flash converter. The last stages have been scaled to save power. Most of the area is used by the capacitors.

With a large differential swing of the input signal ($2\,V_{pp}$) the capacitors can remain relatively small, causing less problems with area, power, etc. The input switches can either be complementary or bootstrapped, see Sect. 3.3.4 and are of minimum length. A potential distortion issue is formed by the bondpad that is connected to the input terminals. Its protection measures must be examined as too much parasitic and non-linear capacitance will affect the performance. Special RF-bondpads are a solution.

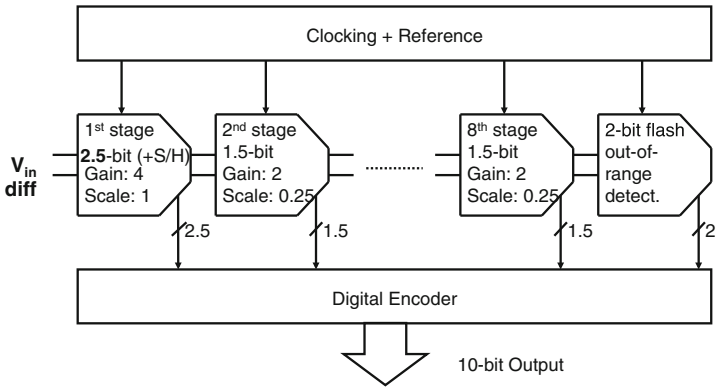
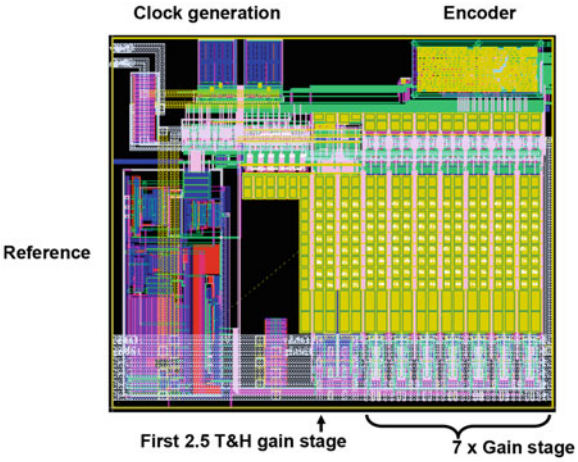


Fig. 8.83 The architecture of a 10-bit pipeline converter. The first stage contains 2.5 bit, the seven remaining stages 1.5 bit [231]

Fig. 8.84 A 10-bit pipeline layout. The first stage contains 2.5 bit, the seven remaining stages 1.5 bit. The depicted area in 90 nm CMOS is $500 \times 600\,\mu m^2$. Courtesy: G. Geelen NXP [231]



Power efficiency is crucial for the application of converters in large system chips. Next to optimizing the power consumption of the individual blocks and techniques such as opamp sharing, also converters are published with adjustable power and performance specifications, see Fig. 8.85 [231].

In a BiCMOS and a CMOS technology, two examples of the potential of this technique are shown [219, 237]. In [219] a 3-bit front-end MDAC is used in a 0.35 µm BiCMOS technology. The uncalibrated performance allows a DNL of 0.2 LSB and 11.2 ENOB with input frequencies close to the Nyquist rate. The sample speed is 125 Ms/s with 1.85 W. In [237] the 3-bit input stage used a capacitor of 1.5 pF, the second 3-bit stage uses 0.4 pF, while the remaining stages have 0.1 pF capacitors. In 65-nm CMOS the sample rate is 1 Gs/s and the performance at 140 MHz input frequency reaches 11.2 ENOB for 1.2 W.

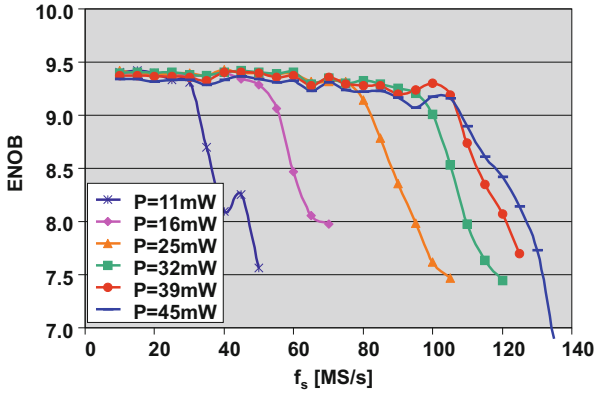


Fig. 8.85 Measured effective number of bits versus sample rate at different power levels. The signal frequency is each time just below $f_s/2$. Courtesy: G. Geelen NXP [231]

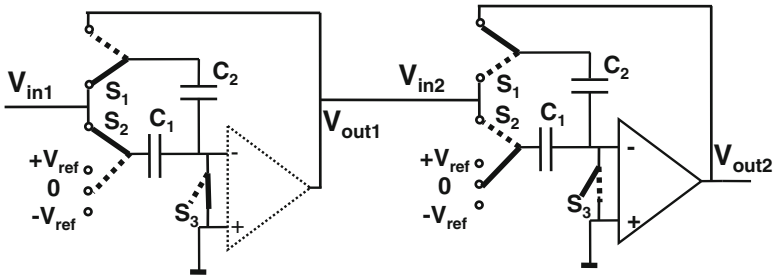


Fig. 8.86 The opamp is shared between two stages in a 1.5-bit pipeline converter. The *left stage* is sampling and the *right stage* is amplifying. During the next half phase the opamp is moved to the first stage which amplifies, while the second stage samples without an opamp. The switches for reconnecting the opamp have been left out for clarity

8.5.4 Pipeline Variants: Opamp Sharing

Power efficiency is a driving force in 1.5 bit pipeline converters. One possibility to improve power efficiency is found in a better use of the hardware. A feature of the configuration in Fig. 8.79 (right) is the fact that during sampling the opamp is redundant. Opamp sharing is a technique that effectively uses the opamp redundancy in the scheme of Fig. 8.86. The opamp now serves two sections of the pipeline converter and reduces the required number of opamps by a factor two [228, 238]. During sampling the input capacitor C_1 and the feedback capacitor C_2 do not need the opamp as long as a separate ground switch S_3 is present. The opamp can be used for the second stage, where the subtraction and multiplication by a factor two takes place. The odd and even sections now run half a sample phase delayed and one opamp serves two sections. The opamp can show a "memory-effect" due to signal charges stored on the internal capacitors. This interference between one sample and

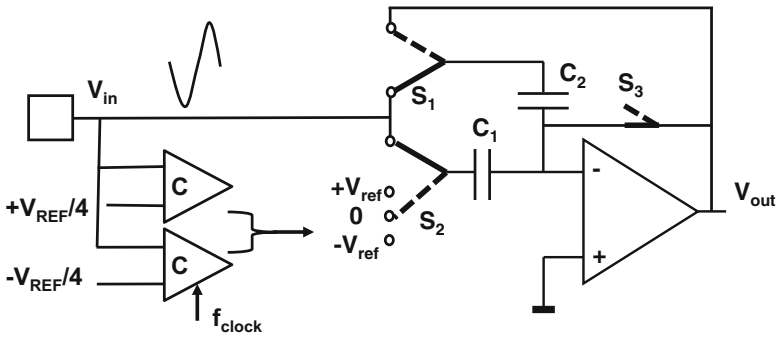


Fig. 8.87 The time-continuous input signal directly fed into the MDAC [240]

the next is undesired. A fraction of the clock period or special switching schemes are used to avoid coupling between the samples.

Some further optimization is suggested in [239]. The capacitor sharing technique reuses the charge formed in the first MDAC for use in the next MDAC. The loading of sample capacitors is avoided and saves power.

8.5.5 Pipeline Variants: S&H-Less

Some designers propose to remove the dedicated front-end track-and-hold function [219, 237, 240]. The pipeline stage in Fig. 8.87 samples now directly the time-continuous input and performs the first processing. Removing the input T&H saves in some cases 25 % of the power, but requires to rebalance the timing and shortens the opamp settling period as the comparators need to decide and settle before the charge is transferred into C_2.

The timing mismatch between the sampling and the activation of the first set of comparators must be addressed as depicted in Fig. 8.88. The problem resembles the math for jitter in Sect. 2.6.1 and random timing skew in Sect. 9.3.3. The timing error τ_d makes that the sample is taken at a different time moment from the comparator decision. The resulting amplitude error $V_{in,d}$ is the result of the signal slope and the delay. This error is multiplied by the MDAC gain and must remain within the reference voltages. While in timing skew and jitter problems the error must remain below 1 LSB, now the safety margin of a multi-bit MDAC is available, e.g. $A \times V_{in,d} < V_{ref}/8$.

Another set of problems appears if the input directly is connected to the input pin of an IC or a source with limited drive capability. The kick-back of comparators and the charge consumption by the MDAC will lead to undesired effects, such as distortion.

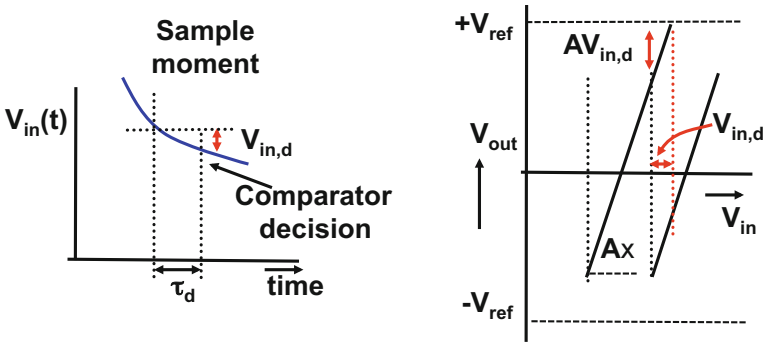


Fig. 8.88 The time-continuous input signal and the comparator decision moments differ

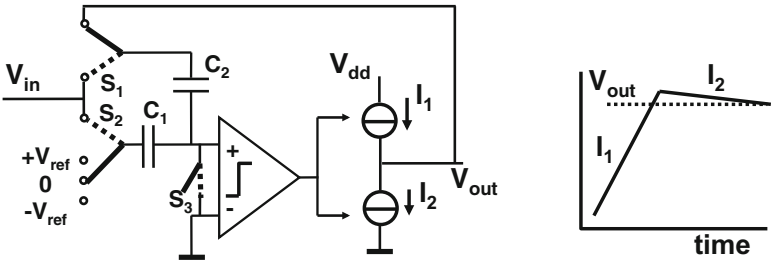


Fig. 8.89 The opamp is removed and replaced by a comparator and two current sources

In [219, 237] 14-bit analog-to-digital converters are presented without an input sample and hold. Designs in a BiCMOS and 65-nm CMOS are compared, see also Sect. 8.5.3. The input buffer for a sample-and-hold less input stage is now a limiting factor. In BiCMOS a strong buffer can be built, but in CMOS calibration of the kick-back is required to improve the performance.

8.5.6 Pipeline Variants: Remove Opamp

Some pipeline converters use digital calibration techniques to overcome analog imperfections. Many authors propose to extend these calibration techniques to eliminate the energy consuming opamps or allow them to perform less.

Replacing the opamps with less performing building blocks or allowing incomplete settling [229, 230] requires more extensive calibration. Sensitivity to changing environmental conditions (power supply, bias, temperature) is unclear.

A radical idea to avoid the opamp in the processing of a pipeline converter is proposed in [241, 242]. A comparator switches on and off two current sources in Fig. 8.89. I_1 is used for fast charging, however, some overshoot must be expected due to the delay between the crossing of the levels at the input of the comparator

and the current switching. A second current source I_2 discharges at a slower rate and reaches the required output level. A higher jitter sensitivity seems likely as the sampled charge is compensated by a charge formed by integrating currents over a period of time: $Q = I_1 \times T_1 - I_2 \times T_2$.

8.6 Successive Approximation Converters

8.6.1 A Simple View

The origins of the successive approximation algorithm certainly include the old-fashioned two-arm beam balance with two equal arms and a pan hanging from each arm. An unknown quantity is placed on one pan and the other pan is filled with known weights till the beam reaches a horizontal position. These weights form an exponential sequence that allows to create the counterweight of the desired resolution.

The successive approximation converter is the electronic equivalent of this beam balance, where a flash converter needs a single clock edge and a linear converter 2^N clock cycles for a linear approximation of the signal, the successive approximation converter (SAR stands for successive approximation register) will convert the signal in N cycles. Each cycle is used to place another (electronic) weight on the pan and check for the result.

Figure 8.90 shows an abstract flow diagram of a successive approximation algorithm. At the start, the output bits a_{N-1} to a_0 are reset to 0. In the first cycle the coefficient corresponding to the highest power a_{N-1} is set to 1 and the digital

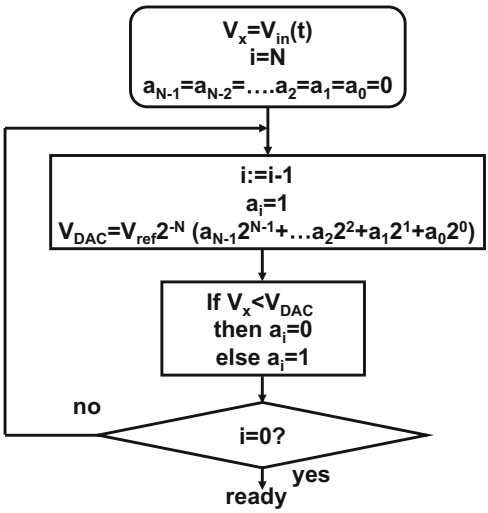


Fig. 8.90 A flow diagram for successive approximation

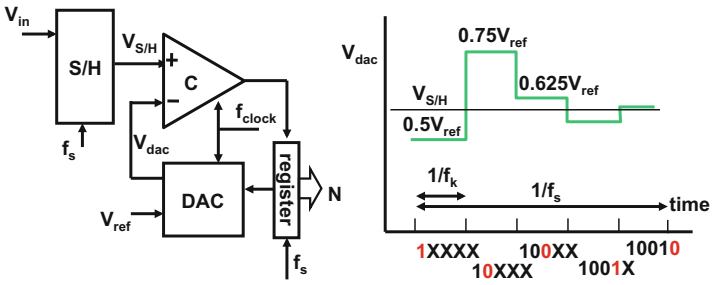


Fig. 8.91 A successive approximation analog-to-digital converter with approximation sequence

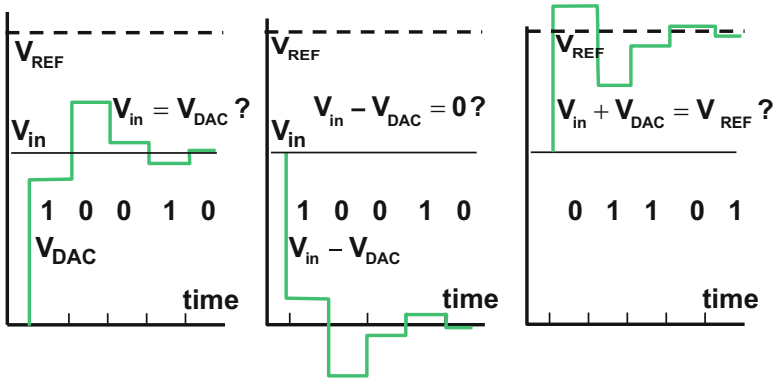


Fig. 8.92 Three forms of successive approximation analog-to-digital converter algorithms: normal, reverse, and complementary

word is converted to a value $V_{DAC} = 0.5V_{REF}$. The input level is compared to this value and depending on the result the bit a_{N-1} is kept or set back to 0. This cycle is repeated for all required bits.

Figure 8.91 shows a circuit implementation. After the signal is stored in the sample-and-hold circuit (or track-and-hold) the conversion cycle starts. In the register the MSB is set to 1 and the remaining bits to 0. The digital-to-analog converter will generate a value representing half of the reference voltage. Now the comparator determines whether the held signal value is over or under the output value of the digital-to-analog converter and keeps or resets the MSB. In the same fashion the next bits in the output register are determined. The internal clock f_{clock} runs at least $N\times$ faster than the sample clock f_s. For every sample, a reset, a sample-and-hold action, and N clock cycles are needed.

In this scheme the digital-to-analog output value approximates the input value. Another implementation reduces the input value by subsequent subtraction.

The approximation algorithm in Figs. 8.91 and 8.92 (left) requires that the comparator operates over the entire input range to full specification. This can be an issue when the signal amplitude comes close to the voltage supply range.

The reverse and complementary forms (Fig. 8.92 (middle and right)) either decrease the signal value to a zero level or complement to reach the full reference. The comparator has a much easier task as it can be optimized for one input level. As can be seen from the example, the processed signal can reach beyond the reference range. In low-power supply circuits this may lead to leakage through forward-biased pn-junctions.

Offsets in the sample-and-hold circuit or the comparator will generate a shift of the conversion range, but this shift is identical for every code. This principle allows sample rates of a few hundred of MegaHertz. The demands on the various constituent parts of this converter are limited. The main problem is a good sample-and-hold circuit that needs a good distortion specification for the entire input range. Next to the sample and hold, the digital-to-analog converter determines the overall linearity and will take up most of the area.

For many applications that do not need the maximum conversion speed that is possible in a technology, successive approximation is a safe and robust conversion principle. In applications with a built-in sample-and-hold function (e.g., a sensor output) the combination with a successive approximation converter is appropriate. In combination with a micro-controller, the register function and the timing can be controlled with software. However, special attention must be paid to processor interrupts that can easily disturb the conversion process.

8.6.2 Charge-Redistribution Conversion

Successive approximation analog-to-digital converters can be built with almost any digital-to-analog converter described in Chap. 7. In the early years successive approximation converters based on resistive structures were dominant [244]. And over the last years designers have rediscovered the potential advantages of resistors at higher frequencies [245], e.g. in designs where the digital-to-analog converter can be shared over multiple converters. Yet, most attention is focused on capacitive structures with their low-power, low-voltage promises.

One of the early fully integrated CMOS successive approximation analog-to-digital converters is known as “charge-redistribution” converter [243, 246]. The principle is shown in Fig. 8.93 and utilizes optimally the properties of CMOS technology: good switches and capacitors. The digital-to-analog converter discussed in Sect. 7.4.1 is now extended with a comparator and some digital control logic. Refer to this section for capacitor implementation details.

In the sampling phase the input signal is stored on a capacitor bank with a total capacitance value of 16C. “C” is the unit capacitor and is laid-out in a standardized manner. The direct use of the capacitors of the digital-to-analog converter to implement the T&H function is one of the strong points of this design. But this feature also creates some application problems as the delivery of charge must be perfectly linear.

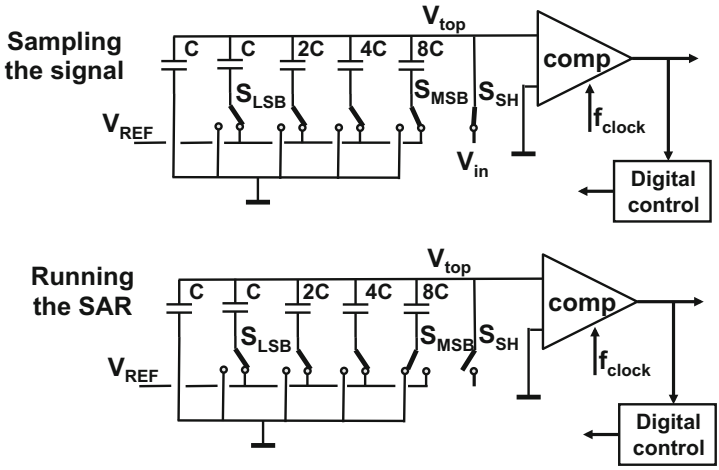


Fig. 8.93 An early implementation of a successive approximation analog-to-digital converter is based on capacitor switching, after [243]

In the second phase the ground plates of the capacitors are switched one after the other from ground to the reference voltage. If the MSB switch is toggled the top plate voltage changes from ground to:

$$V_{top} = V_{in} - \frac{8C}{C + C + 2C + 4C + 8C} V_{ref} \tag{8.43}$$

Depending on the original value of the input voltage, the comparator will decide to keep the MSB switch in this position or return to ground. Every bit is subsequently tested in the same way, thereby minimizing the difference between V_{top} and the comparator reference voltage.

$$V_{top} = V_{in} - \frac{a_3 \times 8C + \ldots + 2^i C}{8C + 4C + 2C + C + C} V_{ref} \tag{8.44}$$

where $i = 3, 2, 1, 0$ in this example and $a_3, a_2, \ldots$ represent the earlier decisions.

The sequence follows the reverse algorithm in Fig. 8.91 (middle). In this implementation the result is a digital code and an output voltage of the digital-to-analog converter that approximates the original input signal. After the last decision is taken, there is a charge balance in place:

$$\sum_{i=0}^{i=N-1} 2^i C V_{in} = (a_{N-1} 2^{N-1} + \ldots.. + a_0 2^0) C V_{ref} + C V_{quant} \tag{8.45}$$

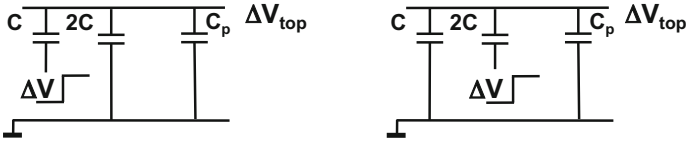


Fig. 8.94 The parasitical capacitor to the top plate attenuates the signal

The input voltage that was stored as charge on the entire capacitive array equals the chosen capacitors times the reference voltage plus some quantization error charge.

In another implementation the sampling is performed on the lower plates of the capacitors and the algorithm will converge to the digital code that complements the input voltage to full-scale. In that case the charge balance is of a complementary form.

In Fig. 8.93 the digital-to-analog converter uses a division of 2^iC on $16C$. For that purpose there is an additional capacitor C on the left side added and it makes understanding of the conversion easier. This capacitor is not necessary. Figure 8.94 analyzes the effect of any capacitor C_p connected to the top plate. A step ΔV on the C capacitor and on the $2C$ capacitor gives

$$\Delta V_{top} = \frac{C}{3C + C_p}\Delta V \qquad \Delta V_{top} = \frac{2C}{3C + C_p}\Delta V \tag{8.46}$$

The same step on a two-time capacitor gives exactly two times amplitude change on the top plate, irrespective of the parasitic capacitance. So the effect of additional capacitive load on the top plate is an attenuation. This property can be used to reduce the swing on the top plate, or scale the residue.

This attenuation becomes prominent in a choice that has to be made: is the input signal stored on the top plate or on the bottom plate in Fig. 8.95? If the signal is sampled on the top plate (Fig. 8.95 (up)), it will experience no attenuation. On the other hand, the swing of the reference voltage will be reduced on the top plate by the capacitive division. This causes (seen from an external perspective) that an input signal with a range $+V_{REF} \leftrightarrow -V_{REF}$ will not fully use the conversion range as it gets attenuated. An advantage is that feeding a signal symmetrical around the switching level of the comparator allows to determine the MSB without using any capacitor bank. In a differential design this feature is very attractive.

Connecting the input voltage to the bottom plate (Fig. 8.95 (bottom)), creates the same path for signal and reference processing. Consequently the amplitude of the input can always span the reference range. Now a common mode voltage must be used to keep charge the top plate to a desired level. This switch and the common-mode voltage must drain the current from sampling. In high performance designs generating this additional voltage may cost power. The advantage for the comparator is obvious: the optimum input voltage can be selected, moreover the effect that the top plate voltage runs out of range is easily prevented.

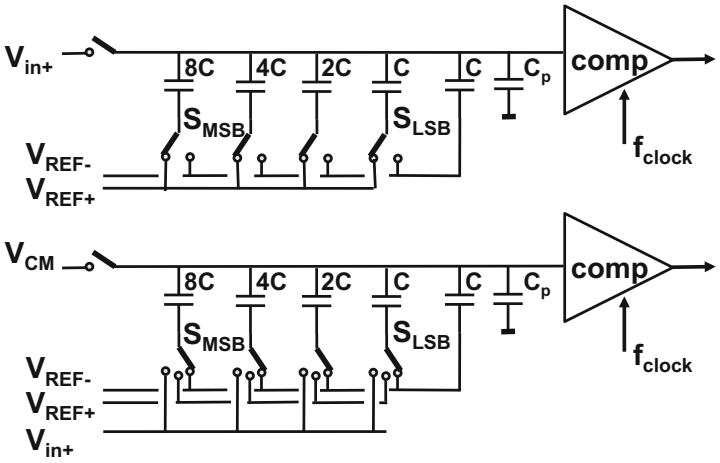


Fig. 8.95 The signal can either be loaded on the top plate or on the bottom plate of the capacitive digital-to-analog converter

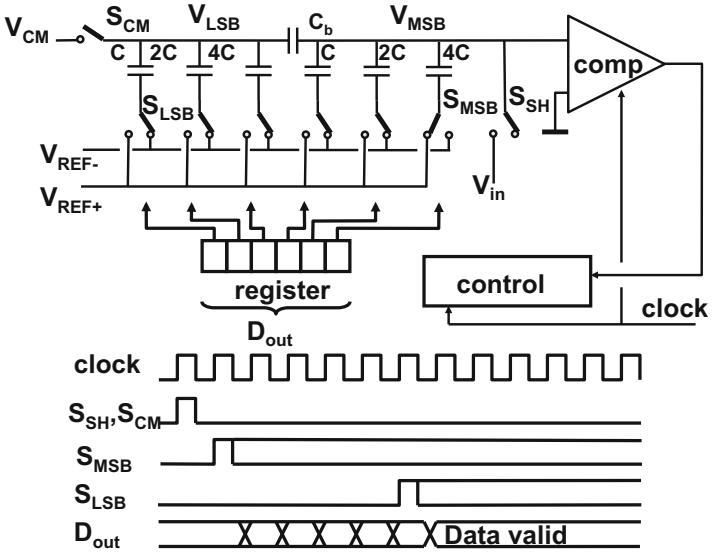


Fig. 8.96 A successive approximation analog-to-digital converter based on capacitor switching with a bridge capacitor [247, 248]

8.6.3 Bridge Capacitor

Figure 8.96 shows a second example of a successive approximation analog-to-digital converter in standard CMOS technology. In a conventional set-up there is

a high-frequency clock $f_{clock} > N \times f_s$ that allows to run N approximation cycles of decisions and capacitive array settings within one sample period. As digital power consumption in CMOS is linearly dependent on the frequency, the digital power of the high-speed logic controller can be considerable. Moreover, the delay between approximation actions is $1/f_{clock}$ and is chosen at a level where the worst-case settling time is always available. In many designs, e.g. [249, 297], the fully synchronous logic is replaced by event driven logic. Now the delay in the operation equals the sum of the real delays, and some speed up is possible.

In the industrial design community, asynchronous logic has been regarded as "tricky business," as it is sensitive to all the physical parameters that a reliable design tries to avoid by choosing the digital abstraction domain. Often proponents of asynchronous logic claim that it presents also advantages with respect to interference. That is a dubious claim. It is true that synchronous logic produces more interference (in the substrate, power supply and other couplings). However, synchronous and asynchronous implementations of the same processor core showed that the interference power of a synchronous core is for 97 % concentrated on multiples of the clock, while the (lower, but still significant) interference power of asynchronous logic is widely spread over frequency bands and is therefore much more difficult to mitigate.[14]

The digital-to-analog converter is implemented as a bridged capacitor array, similar to the architecture in Fig. 7.48. An important difference of the analog-to-digital converter is that the summation node is not a virtual ground node as it is in the digital-to-analog converter. Still, the bridge capacitor allows to use two equally sized capacitor arrays and thereby reduces the overall size. The division of the digital-to-analog converter by means of a bridge capacitor can be explained in various ways. For example, the bridge capacitor loads the MSB side of the array with a selectable fraction of LSB capacitors. Another viewpoint is that the LSB section generates a voltage division on the left side of the bridge capacitor. And readers of the previous sections will recognize a coarse-fine structure, where the size of the bridge capacitor determines the proper connection between the fine range and the MSB steps.

The conversion cycle starts after the sample is loaded on the capacitors through switch S_{SH}. In this example this action also resets the structure, though also a separate reset clock cycle and switches can be used. The reference voltages are chosen with equal but opposite voltages with respect to the signal ground level. The purpose of this successive approximation is to make the signal on the input of the comparator equal to the ground level. To achieve that goal in this implementation the MSB-switch in the sampling phase is connected to the plus reference, while the other switches are connected to the negative reference. After the sampling the MSB can be tested and then sequentially all other bits.

The voltage swing on the input node of the comparator depends on a correct charge sharing between the capacitors connected to the top plates. The bridging

[14]This research was presented at the 2010 ISSCC forum, and still waits for some spare time to write a publication.

capacitor C_b must fit to the LSB capacitor bank with a total value of C_{LSB} and the MSB capacitor bank with C_{MSB}.

If the LSB capacitor of the MSB bank switches, the change in V_{MSB} is

$$\Delta V_{MSB} = \frac{C}{C_{MSB} + \dfrac{C_b C_{LSB}}{C_b + C_{LSB}}} (V_{REF+} - V_{REF-}) \tag{8.47}$$

where C is the unit capacitor. The right-hand term in the denominator represents the series connection of C_b and C_{MSB}.

If the LSB capacitor of the LSB bank switches, the change in V_{MSB} is

$$\Delta V_{MSB,LSB} = \frac{C_b}{C_b + C_{MSB}} \Delta V_{LSB} = \frac{C_b}{C_b + C_{MSB}} \frac{C}{C_{LSB} + \dfrac{C_b C_{MSB}}{C_b + C_{MSB}}} (V_{REF+} - V_{REF-}) \tag{8.48}$$

This change in $V_{MSB.LSB}$ is ideally a $2_{-N_{LSB}}$ fraction of an LSB change in the MSB section. From this equation the bridge capacitor is found as

$$C_b = \frac{C_{LSB}}{2^{N_{LSB}} - 1} = C \tag{8.49}$$

In case the LSB section contains $2^{N_{LSB}} - 1$ unit capacitors. Note that the MSB section drops out of the equation. Any parasitic capacitance on the MSB side will equally attenuate signals from the MSB and LSB sections. Consequently it is advisable the bridge capacitor with its parasitic side to the MSB section (unless this size is polluted with substrate noise).

Parasitic capacitors or additional capacitors on the LSB side will attenuate the LSB steps and a similar errors as in a coarse-fine converter will appear. See Figs. 8.97 and 8.98. In case of parasitic capacitors on the LSB side, the bridging capacitor can be increased.. The bridging capacitor can be rounded to unity if N_{LSB} is large [247], thereby avoiding the need for a fraction of a capacitor with various

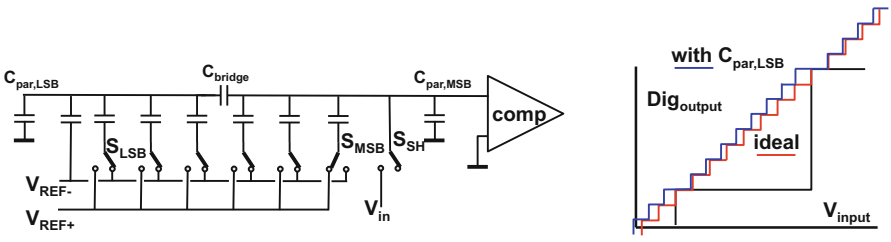


Fig. 8.97 The LSB side of this bridge capacitor array suffers from a parasitic load $C_{par,LSB}$. Now the LSB-steps are too small and a DNL error at the MSB transition will occur

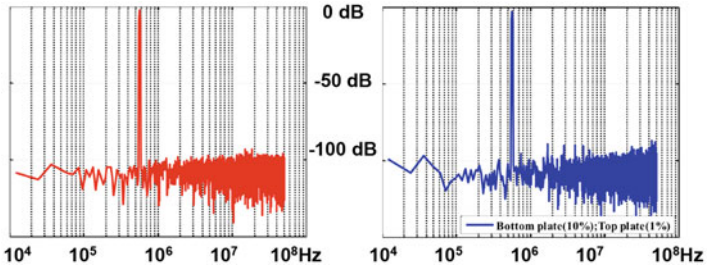


Fig. 8.98 Simulation of a spectrum of an ideal 12-bit successive approximation converter with two arrays of 6-bits (*left*). And with a 1 % parasitic capacitor on the LSB side (*right*). Courtesy: Charles Perumal and Ivan O'Connell, MCCI, Cork

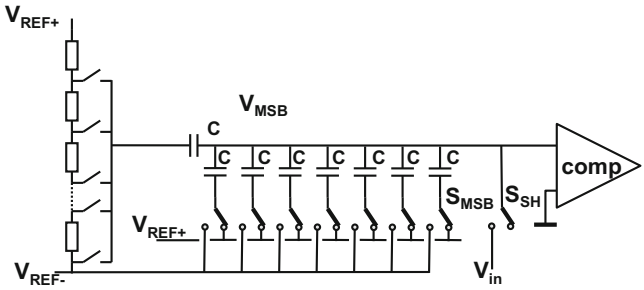


Fig. 8.99 In this successive approximation converter the LSB size is implemented with a resistor ladder

mismatching issues. The bridging capacitor concept can be extended over more sections. Chen et al. [250] uses three sections.

A variant with similar properties as the bridge capacitor design uses a resistor network for the LSB section (Fig. 8.99). This hybrid solution allows a better coarse-fine connection, but requires static current.

Both the resistor string and the capacitor array have been implemented as unary digital-to-analog converters, thereby improving DNL. A unary capacitor array has a similar size for the same kT/C noise as a binary, only the control requires more array, e.g. [251]. The circuit can be operated in a linear search mode (testing every capacitor separately) or in binary search. Then the MSB is tested by connecting four capacitors, and if rejected, the procedure is repeated with two. A form of data-weighted averaging is also applicable.

8.6.4 Speed, Noise, and Kick-Back

The conversion speed is determined by the settling of the digital-to-analog converter and the decision making of the comparator. Especially in larger structures with a lot

of elements this settling time constant τ_{DAC} can be rather long. For reaching 0.5-LSB accuracy in an N bit converter the settling time requirement is

$$t_{settle} > \tau_{DAC} \ln(2^{N+1}) \tag{8.50}$$

The main elements in the settling time constant are the switch resistance and the comparator settling. Included in the switch resistance is the wiring and the limited output impedance of the reference supplies. Often the switches are scaled for the capacitance they drive. In this way the settling of each unit can be kept largely on an equal level. In some designs groups of unit cells forming part of a weight are each driven by their own switches. With 1–10 pF total array capacitance for 10-bit and more resolution, and for sample rates of 100 Ms/s or more, the switch resistance must be budgeted at the 100 Ω level. Harpe et al. [252] discusses implementation details for medium speed converters.

Special attention is needed for the sampling process and the distribution of the input clock to the sampling switches. Bringing a differential clock signal from a generator to the switch that determines the sample moment and also determines the jitter noise is often underestimated. Bringing a low-jitter clock to the decision switches requires area and power consuming buffers in combination with careful lay-out of the PCB and the on-chip wiring.[15]

The comparator speed depends largely on the chosen topology, the current and the technology. As fixed comparator offsets in the successive approximation converter show up as a tolerable signal offset, and do not affect the DNL, the transistor sizes need not to fulfill strict matching conditions. The use of dynamic comparators however does lead to strong charge exchange and kick-back.

Kick-back charge from the comparator can easily disrupt the correct determination of the comparator decision. In Fig. 8.100 this effect is modeled with a feedback capacitor. The comparator decision will induce charge on the top plate that can affect the decision that is being processed or the next decision. Therefore many designers

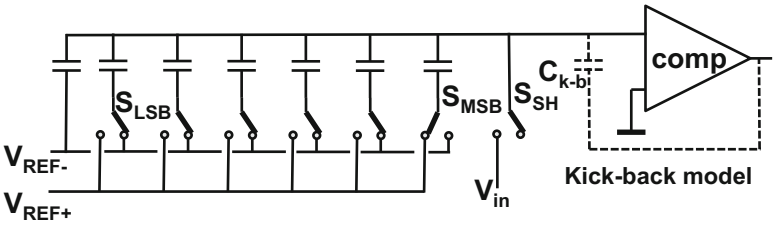


Fig. 8.100 The delicate charge balance in a successive approximation converter based on a capacitive array can easily be disturbed by kick-back from the comparator

[15] Don't be surprised to see 10 mA flow into this part of the circuit. This aspect is hardly mentioned in F.o.M boosting designs.

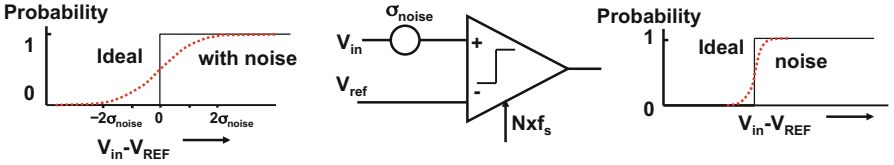


Fig. 8.101 Comparator noise is a major source of performance loss in successive approximation converters. Majority voting (*right*) reduces the effect

avoid simple comparators, as the Strongarm comparator, but use preamplifiers. Differential designs have the advantage that common-mode kick-back is cancelled.

The two major noise sources in a successive approximation converter are the kT/C noise of the sampling by the capacitor array, and the comparator noise. With a large input transconductance the effective value of the input-referred noise is given in Eq. 8.9. In a bandwidth of 1 GHz and with an input transconductance of 1 mA/V the value for the input referred rms noise is $v_{in,noise} = 126\,\mu\text{V}$. For an V_{LSB} size of 0.44 mV the thermal power equals the quantization power, or the converter loses 0.5 ENOB. Taking multiple decisions and using majority voting in Fig. 8.101, the noise is reduced to a level approximated by

$$v_{in,noise} = \sqrt{4kT\frac{BW}{N_{votes}g_m}} \tag{8.51}$$

In [253] majority voting is selectively applied. A helper circuit detect if the comparator is slow and consequently sees a low noise-corrupted signal. In that case a majority voting is applied. This design uses two capacitors of each 9 pF for differential sampling and achieves 11.2 ENOB in 16 kHz bandwidth with 32 ks/s and $V_{DD} = 0.8\,\text{V}$ and $0.3\,\mu\text{W}$ power.

Note that during a successive approximation sequence the comparator will see at most one occasion where the input signal is so close to the top-plate voltage that a BER situation can exist.

8.6.5 Accuracy and Redundancy

In successive approximation converters there are two major error sources around the core process: capacitor mismatch and comparator decision errors. The capacitor array is the frame of reference and should have sufficient quality. In case one or more of the capacitors deviate from their value, INL (and DNL) errors are difficult to avoid, see Fig. 8.102.

An example of present capacitor accuracy is shown in Fig. 8.103. The uncalibrated INL of this 18-bit successive approximation converter reaches a 13–14-bit level. Moreover, the INL pattern differs for every part, indicating there is no

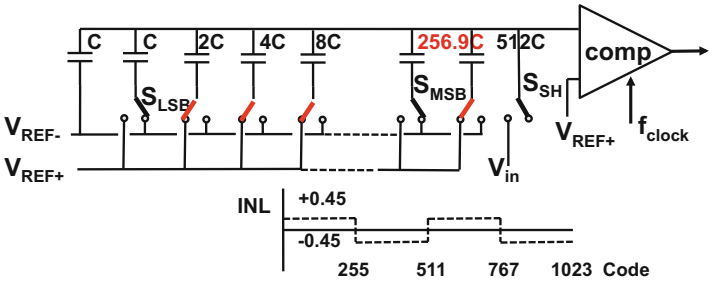


Fig. 8.102 Capacitor mismatch causes INL errors

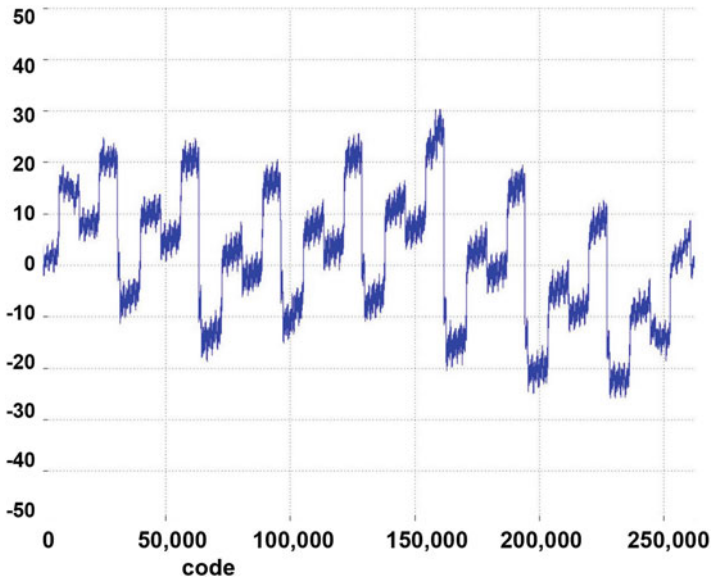


Fig. 8.103 The uncalibrated INL performance of an 18-bit successive approximation analog-to-digital converter [255]. Note that the vertical axis is in 18-bit LSBs! This plot indicates that the design has an inherent capacitor accuracy of 13–14 bit. Note the regularity in the pattern, identifying the largest mismatch in the MSB-3 capacitor in this part. Courtesy figure and private communication: M. Barry, C. Lyden and A. Bannon, ADI, Cork

performance lost due to lay-out inaccuracies. The remaining variation is due to the granularity of the interface in the MIM capacitors in a 0.18 µm CMOS process. A calibration procedure measures the errors and stores the values in fuses. The errors are corrected digitally. Due to the good basic accuracy, the calibration is limited and stable with temperature and aging. This design achieves an integral linearity of 2 ppm (0.5 LSB) after calibration and dither at a sample rate of 5 Ms/s [255].

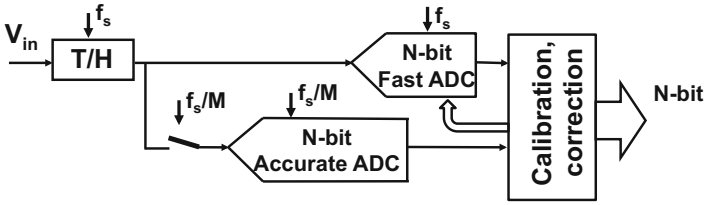


Fig. 8.104 Calibrating a fast analog-to-digital converter by means of a slow but accurate converter

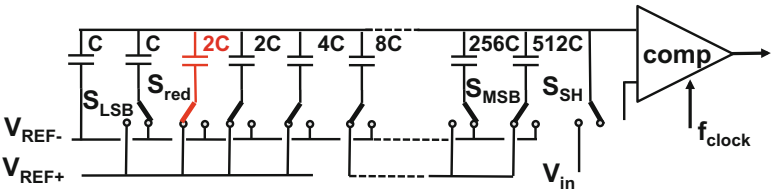


Fig. 8.105 A standard successive approximation circuit is extended with an additional correction stage of 2 LSB

Another form of capacitor correction is to implement the highest capacitor values with some correction capacitors. For example, a capacitor of a nominal weight of 512C is implemented with 510 unit capacitors and 4 selectable unit capacitors to fix errors.[16]

The implementation of repairing capacitor-array errors is shown in Fig. 8.104. The idea is to have a second analog-to-digital converter in parallel to the fast converter. The second accurate converter runs much slower and can be designed as a slow linear approximation converter. It will convert every M-th sample and the digital result is compared to the output of the fast converter. In case of errors, the feedback path will correct the fast converter. This very well-known system is applicable in many converter topologies, and also for successive approximation converters, e.g. [254].

Successive approximation converters can easily be equipped with forms of redundancy for correction of comparator related errors. A very simple mechanism is shown in Fig. 8.105. The standard capacitor array now has a double LSB+1 section (2C) and therefore requires one more clock and decision cycle. In case the incomplete capacitor array settling or the comparator led somewhere to an error of that magnitude, the missing amplitude still can be retrieved. Now the charge balance is

[16] You better do a good lay-out!

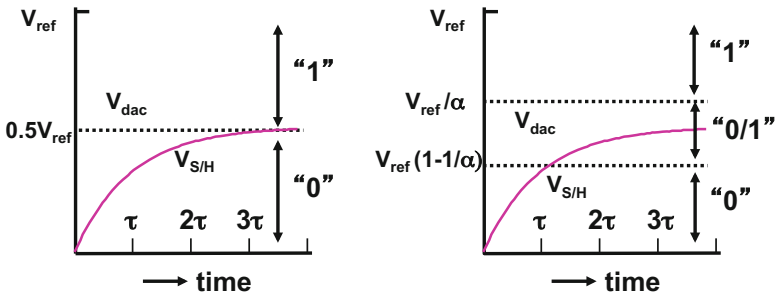


Fig. 8.106 In a base-2 successive approximation the settling of the input signal must reach the final accuracy level before the next step is taken. In non-binary search the base is smaller than 2 and an intermediate region of signal levels exists, where an initial decision can be corrected

$$\sum_{i=0}^{i=N} 2^i C V_{in} = (a_{N-1}2^{N-1} + \ldots.. + a_1 2^1 + a_{red}2^1 + a_0 2^0) C (V_{REF+} - V_{REF-}) + C V_{quant} \tag{8.52}$$

so the logic behind the converter needs to calculate[17]:

$$\frac{V_{in}}{V_{REF+} - V_{REF-}} \Leftrightarrow \frac{(a_{N-1}2^{N-1} + \ldots.. + a_1 2^1 + a_{red}2^1 + a_0 2^0)}{1026} \tag{8.53}$$

Choosing the position of the correction section depends on the expected errors. Errors after the correction section was invoked will not be corrected. But a too small correction section will not grab all errors.

In a successive approximation with a base of 2, the first decision is the final decision on the MSB. It is necessary to let the signal from the track-and-hold stage settle till sufficient accuracy is reached before the decision is taken, see Fig. 8.106. In [251] the base for the digital calculation is not 2 but, e.g., $\alpha = 1.85$. After the MSB decision, the range for the remaining search is not the half of the original range but a fraction $1/\alpha, \alpha < 2$ larger than half of the range. This decision range for a "0," see the right side of Fig. 8.106, extends over the "1" decision level and therefore allows an intermediate range where the initial decision can be corrected. Of course some more clock cycles are needed for this redundancy, however as the signal needs far less settling time the frequency can be a factor 2–3 higher. This type of search is also called: "non-binary" search, compare Sect. 8.4.4.

Figure 8.107 shows the realization of a reduced radix. Here a radix or base of 1.85 is used, that is easily constructed:

For a full 10-bit range it is important to round the unit numbers to such values that the total units add up to 1023. Again the correctness of this sequence follows from the charge balance:

[17] There is no equal sign between both formulas as they differ for the quantization error.

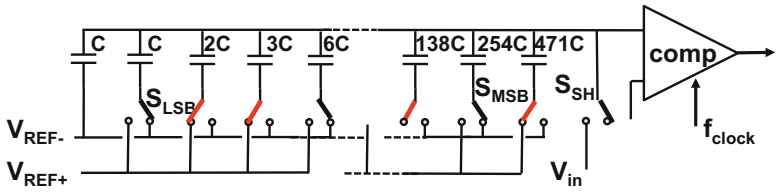


Fig. 8.107 A standard successive approximation circuit (*upper*) is extended with an additional correction stage (*middle*) or with a non-binary sequence

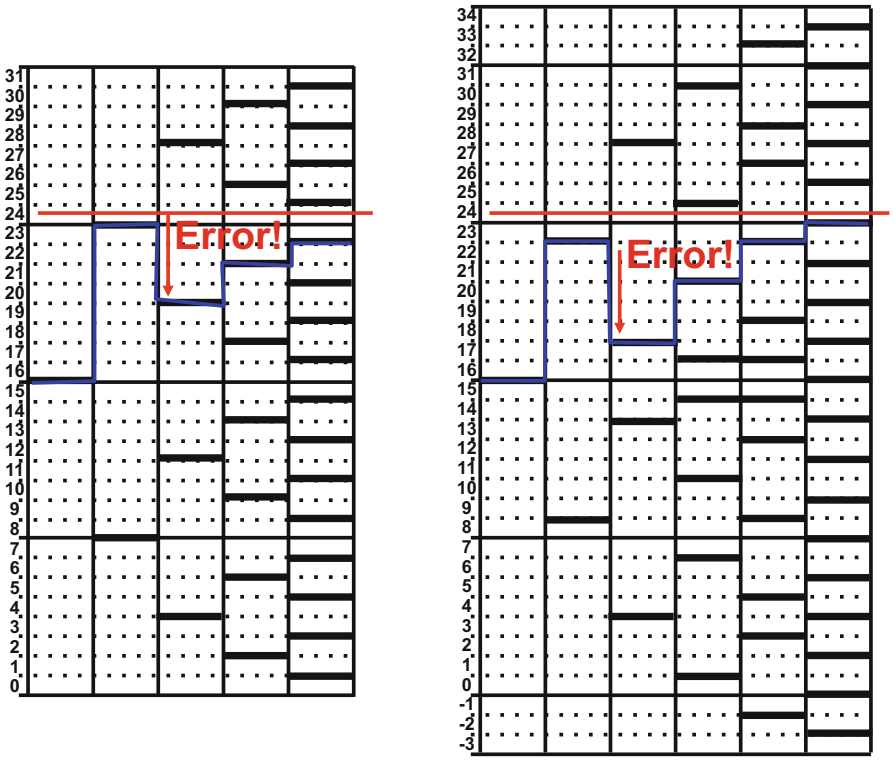


Fig. 8.108 *Left*: In a base-2 conversion an erroneous decision by the comparator in the second cycle cannot be fully recovered. A base-1.85 sequence takes one more clock-cycle but can recover from the comparator error

$$\frac{V_{in}}{V_{REF+} - V_{REF-}} \Leftrightarrow \frac{(a_{10}471 + a_9254 + a_8138 \ldots\ldots + a_36 + a_23 + a_12 + a_01)}{1023} \tag{8.54}$$

Figure 8.108 shows the decision flow of an error in a base-2 and base-1.85 sequence.

In [251] a 10-bit converter is built from a 256-capacitor unary array with 2-bit binary weighted cells. The above-mentioned base-1.85 series of capacitors is

Table 8.5 This table shows the values for 1.85^{Bit} and below a rounded version totaling 1023

Bit	0	1	2	3	4	5	6	7	8	9	10
1.85^{bit}	1	1.85	3.42	6.33	11.71	21.67	40.09	74.17	137.21	253.83	469.59
Rounded	1	2	3	6	12	22	40	74	138	254	471

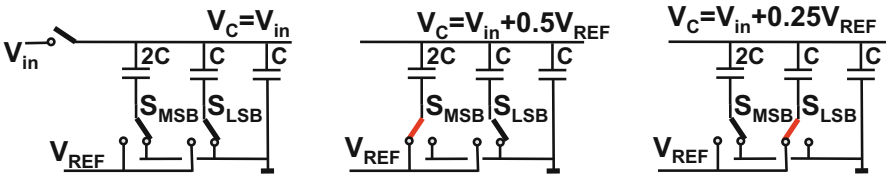


Fig. 8.109 Three cycles in a 2-bit successive approximation: sample, test MSB, test LSB. Example taken from [249]

now implemented in the digital domain. If a new digital-to-analog value must be generated, the pointer in the array is moved by the number in Table 8.5 in Sect. 8.6.5. The same idea resulted in [256] in an 85–90 dB THD performance for 40 Ms/s and 66 mW.

8.6.6 Energy

These successive approximation converters use a limited amount of hardware and good energy efficiencies have been reported [247, 248, 257].

Still the search for lower energy consumption in the capacitor array is continuing. The reason is not so much lowering the energy consumption of the converter, but the energy needed to generate the reference levels. A reference circuit with sufficient performance in terms of settling requires much more power than the successive approximation converter uses.

Figure 8.109 shows the standard operating sequence of a 2-bit successive approximation capacitive array. After the sample has been stored on the top plate, the MSB switch is connected to V_{REF}. The MSB capacitor is charged and the top plate moves to $V_{in} + 0.5V_{REF}$. The reference source delivers the charge change on the MSB capacitor: $Q_{MSB} = 2C \times 0.5V_{REF}$. And the energy needed [249] is $E_{MSB} = CV_{REF}^2$.

Now the MSB test is rejected (as in this example the input voltage was too low) and the LSB is tested. If the MSB is set back and the LSB is switched to V_{REF}, as in Fig. 8.109 (right), the LSB capacitor voltage changes from $0.5V_{REF}$ to $-0.75V_{REF}$. And consequently the energy needed is $E_{LSB} = 1.25CV_{REF}^2$. Even more than the MSB energy! If, on the other hand, the situation of Fig. 8.109 (right) is reached after first discharging the MSB capacitor to ground and after that switching the LSB capacitor, the energy needed is $E_{LSB} = 0.75CV_{REF}^2$. The subtle difference is that if

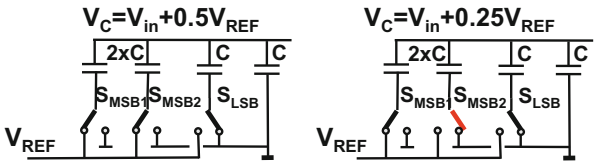


Fig. 8.110 The split capacitor mechanism [249]

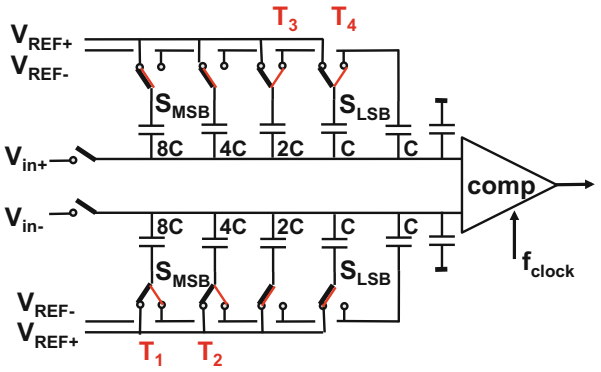


Fig. 8.111 The set-and-down algorithm [164]

both capacitors are simultaneously switched a part of the MSB discharge current is delivered by the reference instead of by the ground.

In Fig. 8.110 a method is proposed to reduce the energy further: split capacitor [249]. On the left-hand side the MSB test is performed. The MSB capacitor is a parallel connection of two LSB capacitors in this 2-bit example. After the comparator decides to reject, only one of the branches of the MSB capacitor is switched to ground. The top plate voltage drops to $V_{in} + 0.25V_{REF}$ and consequently the reference source must deliver the charge change of the remaining MSB branch. As this charge change is only $Q_{MSB,left} = 0.25CV_{REF}$, the energy delivered for activating the LSB test is $E_{LSB} = 0.25CV_{REF}^2$, which gives a five times reduction of energy consumption.

After the initial publications had shown real energy saving potential, a large number of proposals have been made. Compared to the worst case, some (complex) energy saving schemes spent only 10 % of the energy. A proposal with several interesting features is the "set-and-down" sequence [164]. The topology of Fig. 8.111 is fully differential. During the reset and sampling phase the bottom plates of all capacitors are charged to the positive reference voltage. This is done with slow PMOS devices, but needs to be done only once per cycle. After the input signal is sampled, directly the sign and MSB is determined. The capacitive DAC is used for the MSB-1 and lower. If not limited by kT/C noise, this reduces the capacitor array by half. Now the sequence starts. At T_1 the MSB-1 is tested and kept. Only the lower array is used. At T_2 the MSB-2 is tested, but is rejected. In the original

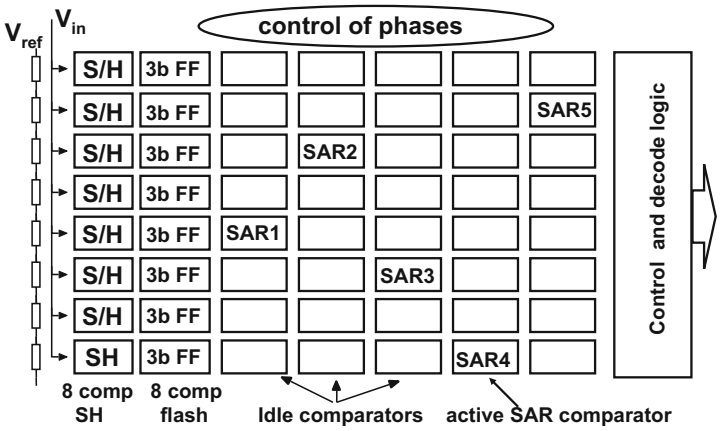


Fig. 8.112 Multi-step analog-to-digital converter, design: J. v. Rens

sequence, now the charge situation must be brought back, but as this is a differential design, another option exists: leave the error bit and continue on the other side to test at time T_3 and T_4. The observation is that the differential nature of this scheme produces $+1, -1$ instead of $1, 0$. As the array during operation is only discharged by NMOS devices, the speed of operation can go up.

8.6.7 Alternative Successive-Approximation Analog-to-Digital Converter

Successive approximation converters have maximally utilized the benefits of CMOS scaling. By doing so, there is a danger that people believe that a charge-based converter is the only possible topology. This paragraph describes a successive approximation converter, that was designed around 1990. It is based on a resistor ladder and combines various aspects of the architecture in a different manner.

The multistep analog-to-digital converter[18] (Fig. 8.112) is an 8-bit converter based on a technique involving a combination of successive approximation, flash and time-interleaving. The comparator for this design is shown in Fig. 8.26. The three most significant bits of the conversion are determined by means of a flash conversion; the remaining 5 bits are realized through successive approximation. Multiple time-interleaved signal paths have been used to increase the maximum sampling frequency.

The hardware of the analog-to-digital converter consists of an array of 56 comparators with a built-in sample and hold stage. The array is grouped in seven

[18]This section is based on a design by J. v. Rens.

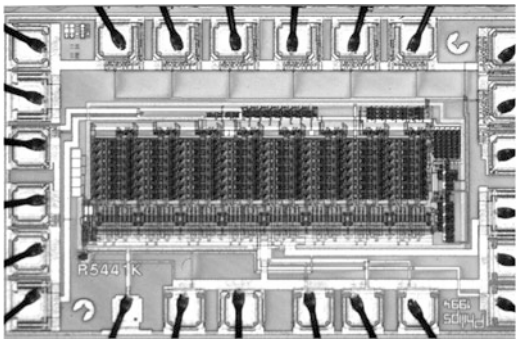


Fig. 8.113 Photograph of multi-step analog-to-digital converter TDA8792

channels of eight comparators in a flash structure. The channels operate in a time-interleaved manner. The actual conversion takes place in 7 clock cycles. First, the input signal is sampled by, and stored in, one channel of eight comparators (sample phase). A flash decision generates the three coarse bits and selects the comparator that stores the replica of the unknown input signal closest to a reference voltage. This comparator is used in a successive approximation loop to determine the remaining bits, while the other comparators are idle.

The input ranges of the eight comparators that form the flash structure determine the signal input range of the analog-to-digital converter. Note that use of parallel signal paths can be successful only if the different channels match well. Offset, gain and timing mismatches between multiple channels give rise to fixed patterns which manifest themselves as spurious harmonic distortion in the frequency domain. The effect of offset is minimized by the use of the previously described offset-compensated comparators. Gain mismatch is minimized by the use of a common resistor ladder digital-to-analog converter and timing mismatch by the use of a master clock which determines the sampling moments of all the channels.

This converter was produced as a stand-alone product TDA8792, see Fig. 8.113, and was the building block for many video integrated circuits.

Based on a single comparator architecture of Sect. 8.1.10, Fig. 8.26 a flash, a single comparator successive approximation and a multi-step pipeline converter were designed.

Table 8.6 shows the main specifications of the three analog-to-digital converters. Basic to all converters is the comparator of Fig. 8.26, in which signal speed and accuracy have been traded off versus power. The decisive factor for the power comparison is the comparator current. In the successive approximation design this current is higher because this comparator has to handle a larger signal span. The lower kick-back of the single comparator in the successive approximation analog-to-digital converter also makes it possible to increase the ladder impedance. All the converters have been extensively used in consumer ICs: digital video, picture-in-picture, instrumentation, etc. In 0.5 μm CMOS the 8-bit multi-step runs at 50 Ms/s, while the 9-bit version achieves 8.2 ENOB.

Table 8.6 Specifications of the three analog-to-digital converters

A/D converter	Flash	Successive app.	Multi-step
Resolution	8 bit	10 bit	8 bit
Sample rate	25 Ms/s	2 Ms/s	30 Ms/s
Differential linearity	0.6 LSB	0.5 LSB	0.5 LSB
Integral linearity	0.6 LSB	1 LSB	0.6 LSB
ENOB at input 4.43 MHz	7.4	8.5 (1 MHz)	7.4
SINAD (4.43 MHz)	46 dB		46 dB
SD (2–5 harm, 4.43 MHz)	>52 dB		>52 dB
Input bandwidth (1 dB)	>70 MHz	20 MHz	70 MHz
Input signal swing	2 V	1.5 V	1.6 V
Ladder resistance	1200 Ω	4800 Ω	1200 Ω
Active area	2.8 mm^2	1.2 mm^2	1.1 mm^2
		0.4 mm^2 (8 bit)	
Technology	0.8–1 μ (1 PS, 2 Al)		
Current	55 mA	3 mA	13 mA
Current/comparator	200 μA	500 μA	200 μA
Number of comparators	256	1	56

All the converters are based on the comparator shown in Fig. 8.26. The analog-to-digital converters in 1, 0.8, 0.6 μm CMOS technology were used on several Philips Semiconductor production chips [258, 259]

8.7 Algorithmic Converters

In the previous examples of successive approximation converters the searching process is implemented by comparing the input value to a set of values from the digital-to-analog converter. Next to the design of the sample-and-hold, the accuracy and the area are determined by the digital-to-analog converter. Algorithmic or "cyclic" analog-to-digital converters keep the reference value constant and avoid a large digital-to-analog structure. By capacitive manipulation the signal is modified [216, 260].

A flow diagram of a basic algorithm is shown in Fig. 8.114. The value V_x is set to the input value and compared to half of the reference. If V_x exceeds half of the reference, this value is subtracted. The remainder is multiplied by 2 and treated as the new input value of the process. This multiplication is an advantage over the elementary successive approximation algorithm. Now errors in the smaller bits count less. Obviously the result of N executions of this flow diagram is

$$V_x = 2^N V_{in}(t) - V_{ref}(a_{N-1}2^{N-1} + \ldots a_1 2^1 + a_0 2^0) \tag{8.55}$$

If the remainder V_x is set to zero (ideally it should be less than an LSB), V_{in} will equal a binary-weighted fraction of the reference voltage. The critical factors in this algorithmic converter are the offsets and the accuracy of the multiplication by 2. The

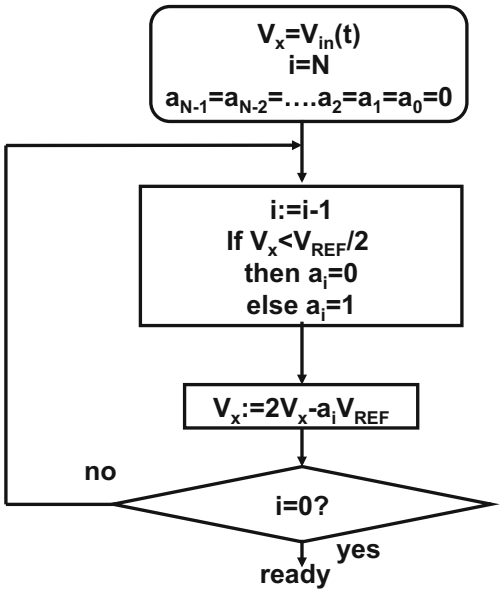


Fig. 8.114 A flow diagram for a cyclic converter

total multiplication error must remain within one LSB. If the amplification equals $(2+\epsilon)$, the difference between the value at the MSB transition and the (MSB-1LSB) transition (the DNL at that code) equals

$$D = (2+\epsilon)^{N-1} - [(2+\epsilon)^{N-2} + (2+\epsilon)^{N-3} + \ldots (2+\epsilon)^1 + (2+\epsilon)^0] \quad (8.56)$$

$$= (2+\epsilon)^{N-1} - \frac{1-(2+\epsilon)^{N-1}}{1-(2+\epsilon)} = \frac{1-\epsilon(2+\epsilon)^{N-1}}{1-\epsilon} \approx 1-\epsilon 2^{N-1}[\text{ in LSB}]$$

The error in the multiplication factor is itself multiplied by the term 2^{N-1}. In order to keep the DNL sufficiently low $\epsilon < 2^{-N}$.

Figure 8.115 shows a basic circuit topology of the converter. After the sample-and-hold circuit has acquired a sample and all capacitors have been discharged, the signal is multiplied by two and compared with the reference voltage to generate the MSB bit. Based on this bit zero or the reference voltage is subtracted from the signal. This remainder signal is fed back to the sample and hold for the next run.

Offsets are critical. The comparator offset must remain under an LSB. In switched-capacitor technique the offsets at the inputs of op-amps and comparators can be removed. The remaining problem is the required accuracy of the multiplication by 2. The minimum capacitor value is mostly determined by the accumulated noise. And the minimum gain of the operational amplifiers is given as in Eq. 8.32. The implicit mismatch of the capacitor structure may jeopardize accuracy. However, a careful lay-out where the capacitor "2C" is built from two parallel capacitors "C" and properly surrounded by dummy structures will reduce this error source to the 10 to 12 bit level. The injection of charge by the switches is especially an issue in older

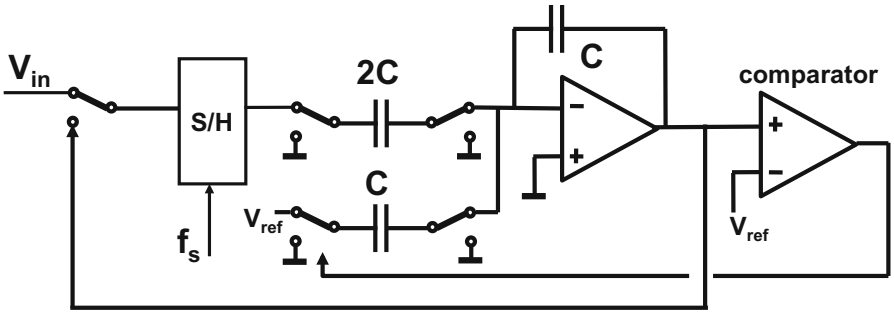


Fig. 8.115 An example of a cyclic analog-to-digital converter, after [220]

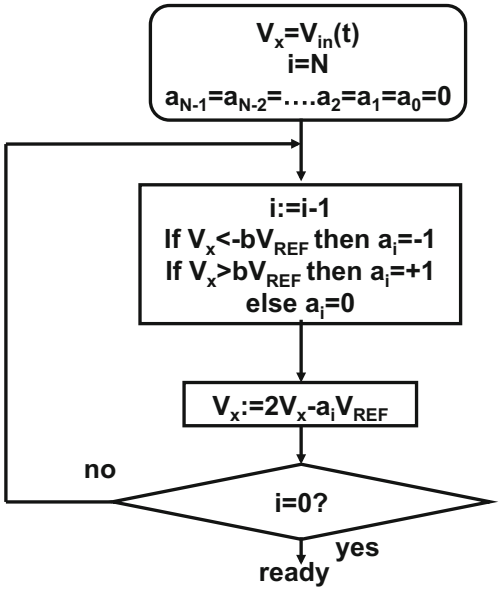


Fig. 8.116 The redundant signed digit algorithm [261]

technologies. The channel charge in the relatively large switch transistors varies with the signal and will affect the overall gain, see Eq. 3.2. Differential operation or special switching sequences help to reduce this effect [220].

Several accuracy issues can be removed if some redundancy is built in. The flow diagram in Fig. 8.116 shows that instead of a single decision now the signal is compared to values bV_{ref} and $-bV_{ref}$, with $b \approx 0.25$. The redundancy is due to the three values that each coefficient can take: $a_i = (-1, 0, +1)$. In simple terms: the algorithm assigns only $a + 1$ or -1 value to a coefficient if the signal is unambiguously positive respectively negative. In case of doubt, the signal is left unaltered. In Fig. 8.117 the algorithm converts a signal $V_{in} = 0.6V_{ref}$. The first decision is therefore $a_{n-1} = +1$, and the remainder is formed as $(2 \times 0.6 - 1)V_{ref} = 0.2V_{ref}$. This new residue value leads to an $a_{n-2} = 0$ decision and the next remainder

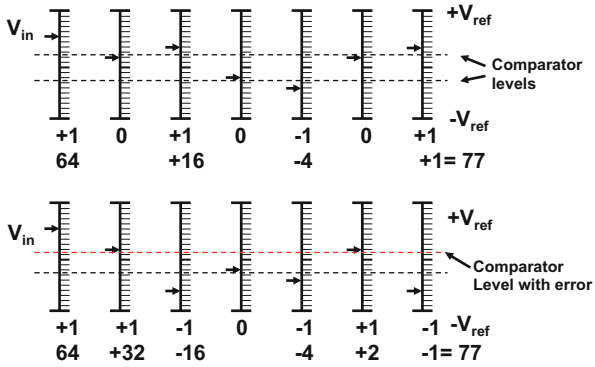


Fig. 8.117 *Upper sequence*: an input voltage of $0.6V_{ref}$ is converted in seven cycles with comparator levels at $\pm 0.25V_{ref}$. *Lower sequence*: the positive comparator level is now at $0.15V_{ref}$ e.g. due to offset, however with the help of redundancy is the same code is reached: $77/128 = 0.602$. Note that the exactly matching end result is partly a coincidence. The difference at the one-but-last stage is 2 LSB

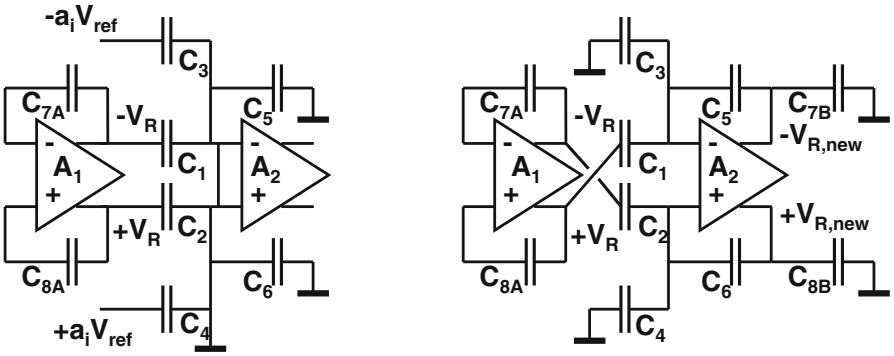


Fig. 8.118 The redundant signed digit (RSD) converter [261]

is simply $2 \times 0.2V_{ref}$. The process repeats until after seven cycles a value of 77 is found: $77/128 = 0.602$.

In the lower part of the figure, the upper comparator level is reduced to $0.15V_{ref}$, e.g. due to offset. Still the algorithm converges to the same overall result. The RSD algorithm[19] creates in this way robustness for comparator inaccuracies. Figure 8.118 shows an implementation of the RSD principle [261]. During the first phase of the cycle the residue is stored in capacitors C_{7A}, C_{8A}. The result of the comparison results in a value for a_i. In the second phase of the cycle, the signal is multiplied by two by means of the cross-coupling of the differential signals on C_1, C_2. This trick

[19]The general form of this principle is in [261] identified as the Sweeney–Robertson–Tocher division principle.

allows that C_1, C_2, C_5, C_6 are all equal. Gain errors due to capacitive mismatches are reduced by interchanging the pair C_1, C_2 with C_5, C_6 for every odd–even cycle. The new values for V_R are stored on C_{7B}, C_{8B} which take the place of C_{7A}, C_{8A} in the next cycle. The typical performance of algorithmic converters is in the 12-bit range at relatively low power consumption. Applications are found in sensors, e.g. [262, 263].

Example 8.13. Compare the 1-bit pipeline converter and the successive approximation converter.

Solution. Both converters use a logarithmic approximation of the signal: first the MSB is decided, and based on the outcome a residue is evaluated. Both principles use a track-and-hold circuit to store the signal. Most bandwidth related specifications will therefore be comparable.

Where the standard successive approximation converter is executing the algorithm in time, the pipeline converter uses additional hardware stages. The pipeline converter can reach a high throughput, because the intermediate results are shifted from one hardware section to the next. The successive approximation converter can achieve a similar throughput rate, when N parallel converters are used.

Mismatch affecting comparators is reduced in the pipeline converter by going to reduced base or 1.5-bit schemes with calibration. In a standard successive approximation converter the comparator offset shows a signal offset and does not affect the transfer. However, multiplexed successive approximation converter also a calibration is required, which is mostly done at an architecture level.

A major difference between both converters is that in all forms of pipeline conversion, an **amplification stage** is used to suppress the errors and noise from the lower bit stages. This feature is missing in successive approximation implementations. The amplification stage is the most power hungry part in a pipeline converter, but it allows to achieve a higher signal-to-noise ratio. The main trade-offs are summarized in Table 8.7.

Table 8.7 Comparison of pipeline and successive approximation

	Pipeline	Successive approximation
Sample cap	First capacitor: kT/C	Total capacitor: kT/C
Comparator offset	Must be removed	Becomes signal offset
Accuracy	In DAC, subtraction and amplification	In DAC
Noise	kT/C and amplifier	kT/C and comparator
Noise scaling	Scales due to amplification	Full noise
Speed	Limited by opamp settling	Limited by DAC settling, comparator
Power	In amplifier	In comparator and CV^2 of DAC
Power efficiency	Good	The best
Speed improvement	By throughput	By time interleave
Main feature	The multiplication-by-2	Direct comparison of signal

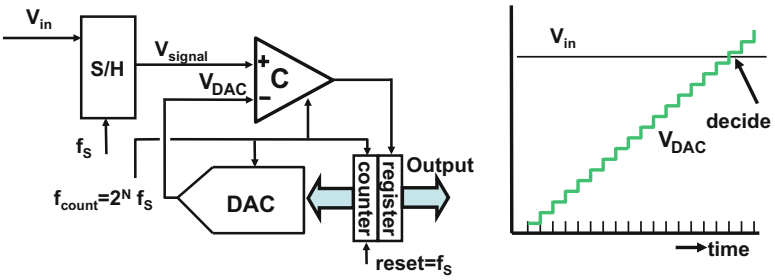


Fig. 8.119 A counting or slope analog-to-digital converter

8.8 Linear Approximation Converters

Flash converters, coarse-fine and pipe-line converters produce at every clock cycle a result. In the case of a flash conversion that will be after a latency time of 1 or 2 clock periods, and for multi-stage conversion that delay may last some $N + 3$ clock periods. Linear approximation methods and converters need for their operation an operating clock frequency that is at least $2^N \times f_s$ higher in frequency than the sample rate. Even if the clock rate runs at the technological maximum, the amount of results per second is limited for linear approximation methods.

Figure 8.119 shows a simple implementation of a counting analog-to-digital converter, also called a "digital-ramp" or "slope" converter. On an edge of the sample pulse the input is sampled and the counter is reset. A high-speed clock will start incrementing the counter. The counter's output is fed to the digital-to-analog converter. If the level of the input sample is reached, the comparator generates a pulse and the register will copy the counter value and deliver this value as the conversion result to the succeeding logic. An example of a linear converter with clock subdivision is described in [264].

These counting or slope converters are often applied in image sensors, Fig. 8.120. The pixel array of an image sensor is read-out via its columns. A row of slope analog-to-digital converters is used to process all pixels of the selected row of pixels. The digital-to-analog converter is shared between all comparators. The simple structure of a counting or slope converter fits well to the narrow pitch of less than 10 μm. The relatively slow speed requirement should allow a medium accuracy of 10–11 bit.

In a densely spaced configuration as in Fig. 8.120, there is little room for large and accurate transistors. The comparators introduce random mismatches between the channels. Often a form of offset correction and $1/f$ noise suppression is desired. In Fig. 8.120 (left) first a reference value, applied to all comparators, is measured and then the signal. In the digital domain the difference is calculated, and most inaccuracies will disappear. Obviously the two sampling events double the quantization power.

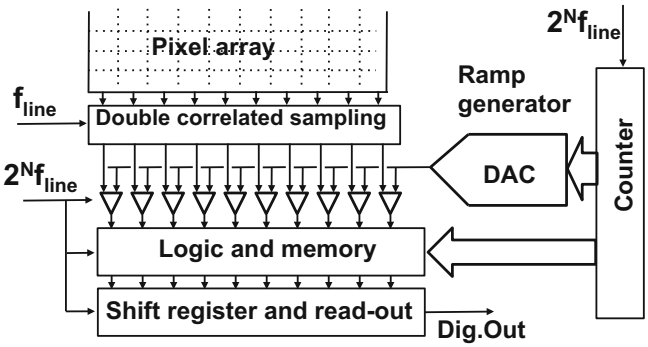


Fig. 8.120 A basic architecture for a slope analog-to-digital converter in an image sensor. f_{line} is the line repetition rate (for simple television quality 15.625 Hz)

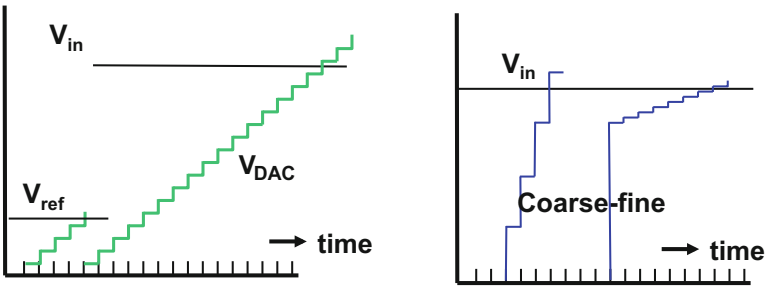


Fig. 8.121 Two clocking schemes for slope analog-to-digital converters. *Left*: First a reference value is converted and then the signal. Subtraction of the two digital result eliminates offsets, $1/f$ noise, etc. *Right*: coarse-fine algorithm: first with coarse steps the value range is determined after which only in that range a fine step is applied [265]

In [265] the counting idea is implemented in a coarse phase with 2^{N-n} increments of size $2^n V_{LSB}$, followed by a fine phase with 2^n steps of V_{LSB} size, thereby reducing the long conversion time, see Fig. 8.121 (right). Similarities to algorithmic converters certainly exist, compare Sect. 8.7.

The counting converter can be turned into a tracking analog-to-digital converter [266] by changing the counter in an up-down counter and connecting the comparator output to the up-down switch, Fig. 8.122. The system will operate in a way that the counter and digital-to-analog converter output will follow the input signal. The acceptable slew-rate of the signal is limited by the speed of the counter clock and the V_{LSB} size. The accuracy is determined by the digital-to-analog converter. Samples can be taken at a rate f_{count}, but also at integer fractions of f_{count}. Yet, with simple means a robust analog-to-digital converter can be built, e.g. for microprocessor interfacing of slow signals. The same type of feedback loop can also be used in other domains, e.g. mechanical tracking systems. Some offset-correction systems, e.g. [267], show similarity with a tracking analog-to-digital converter.

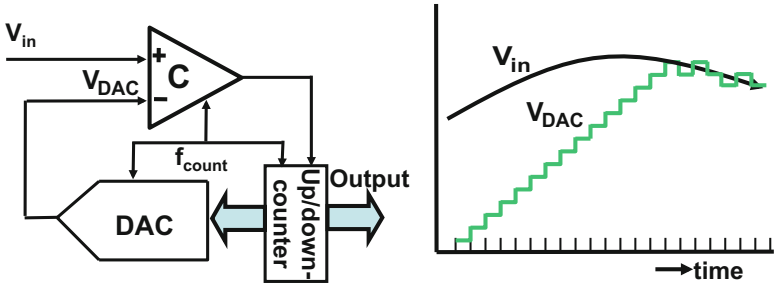


Fig. 8.122 A tracking analog-to-digital converter

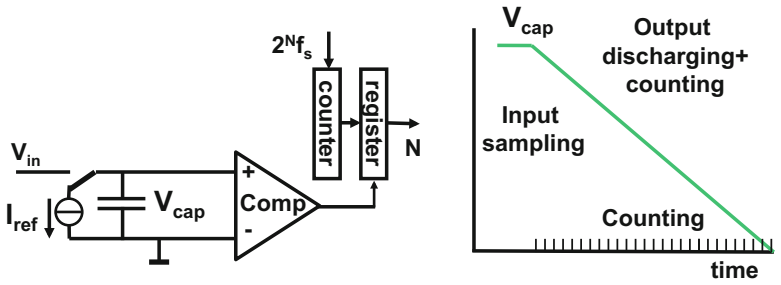


Fig. 8.123 The Wilkinson converter was proposed around 1950

The tracking converter is a "delta-modulator," which is the predecessor of the sigma-delta converter, see Chap. 10. The tracking converter is a feed-back system where the counter acts as an integrator.

The Wilkinson converter starts from the assumption that the signal is sampled on some storage medium, often a capacitor. The conversion starts when a reference current discharges the capacitor value until a threshold set by a comparator is reached. This simple mechanism is rather robust, but the comparator offset is added to the signal (Fig. 8.123).

A "dual-slope" converter is suited for slowly varying input signals. In Fig. 8.124 an integrator circuit integrates the input signal during the fixed sample period. During a second time-frame a reference current discharges the integrator, while a counter measures the number of clock periods needed until the starting level is reached again. The maximum number of clock cycles is around 2×2^N. The input integration has a transfer characteristic comparable to a sample and hold circuit:

$$H(\omega) = \frac{\sin(\pi \omega T_{int})}{\pi \omega T_{int}} \tag{8.57}$$

The low-pass characteristic of this operation makes that this principle is tolerant to RF noise and interference.

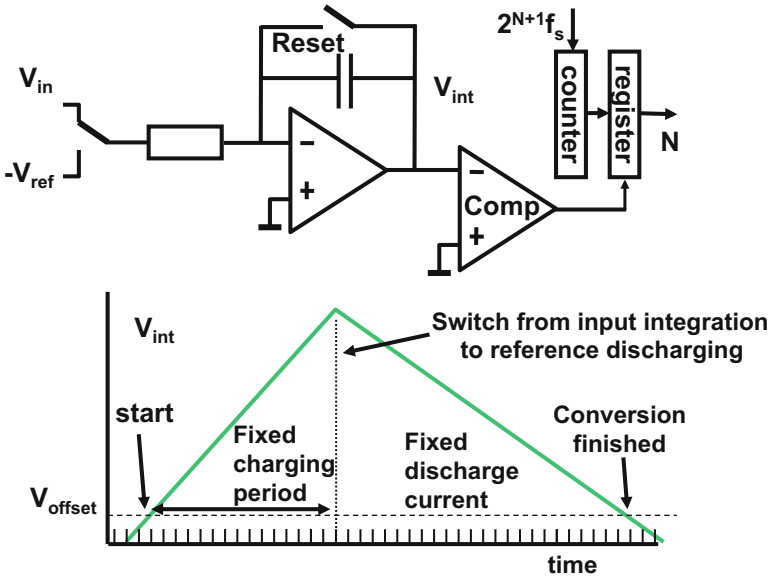


Fig. 8.124 A "dual-slope" analog-to-digital converter

A dual-slope converter is an example of a zero-point detecting method or zero-crossing method. The converter determines the value of the unknown signal by subtracting an equivalent signal from the digital-to-analog converter until the zero starting level is reached. The advantage of zero-crossing methods is that the system needs to be linear only around the zero level. Voltage dependency of the integration elements (non-linear capacitor) in Fig. 8.124 will not impair the conversion accuracy. Moreover an offset in the comparator or integrator is implicitly canceled by the operation as long as the offset in the crossing of the rising edge is still present when the signal returns to zero at the end of the cycle. Hysteresis around the zero crossing or memory effects (e.g., interface trapping) in the capacitor cannot be tolerated.

Dual-slope converters find their application especially in multi-meters and in harsh industrial environments: e.g., operation above 200 °C, oil-drilling, or chemistry.

Example 8.14. Compare at flash converters, successive approximation converters and dual-slope converters with respect to DNL, INL, and absolute accuracy in case of comparator threshold mismatch.

Solution.

Example 8.15. In an IC process input pairs (as used in comparators, gain stages, etc.) suffer from $\sigma_{Vin} = 50\,\text{mV}$ maximum uncorrelated errors, while resistors can be

	DNL	INL	Absolute accuracy	Speed
Flash	Poor, limited by comparator mismatch	Poor, due to ladder and comparator mismatch	Poor, due to ladder and comparator mismatch	highest 10 Gs/s
Successive approximation	Good, limited by digital-to-analog converter	Good, limited by digital-to-analog converter	Poor, due to comparator offset,	Up to 300 Ms/s
Dual-slope	Good, digital-to-analog converter	Good, digital-to-analog converter	Excellent, only drift	1–100 ks/s

Table 8.8 Comparison of converters in case of extreme mismatch

ADC architecture	Remarks	Resolution
Flash converter	2^N parallel comparators, limited by comparator mismatch	N= 4 bit
Pipeline and dual slope converter	Comparator error will be cancelled, INL,DNL determined by digital-to-analog conversion	$N \approx 10$ bit
Logarithmic approximation, successive approximation	Comparator error results in offset, INL,DNL determined by digital-to-analog conversion	$N \approx 10$ bit

made with 0.1 % accuracy. Which ADC architectures can be made advantageously in this process (give indication of the resolution).

Solution. If a signal range of approximately 1 V is assumed, a 5 % error results. The main trade-offs are summarized in Table 8.8.

8.9 Other Conversion Proposals

Many other principles exist to convert signals from the physical domain to bits. Not all of them are relevant for a larger community, yet some of them may be considered in specific circumstances.

8.9.1 Level-Crossing Analog-to-Digital Conversion

In the previous sections analog-to-digital converters were designed by sampling the signal and subsequently quantizing the signal to reference levels. In this process rounding errors occurred, that were labeled quantization errors. The sequence of sampling and quantizing can also be reversed. The level-crossing analog-to-digital

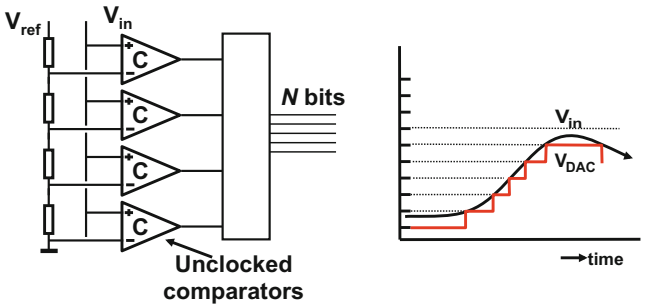


Fig. 8.125 The principle of a level crossing analog-to-digital converter

converter in Fig. 8.125 [268–270] generates a new digital output code at each time moment an amplitude quantization level is passed. In its simplest form, this is a flash converter with non-clocked comparators. With infinite time resolution this level-crossing algorithm will lead to a digital representation of the input signal with some harmonic distortion, depending on the density of the levels. There is no folding back of spectra and a rather high quality signal representation can be obtained. The signal information is coded in the number of levels crossed but also in the timing information that is continuous. In a conventional digital system, it is impractical to process this pulse-width modulation signal and therefore a rounding to a time grid is needed. That step introduces rounding errors and quantization power. Suppose that the rounding is towards a time grid specified by a sample frequency f_s with a time period T_s. If the level crossing occurs ΔT_s before a sample moment nT_s, the amplitude error is

$$\Delta A(nT_s) = \frac{dV_{in}(t)}{dt}\Delta T_s \tag{8.58}$$

Assuming that the signal $V_{in}(t) = A\sin(\omega t)$ is so slow that only one level is passed during a time period T_s, and that the probability of the occurrence of a level crossing moment is uniformly distributed, and that the level crossing is independent of the signal derivative, the expectation value of the error is

$$E((\Delta A)^2) = \left(\frac{dV_{in}(t)}{dt}\right)^2 \times E(\Delta T_s^2) = \frac{A^2\omega^2 T_s^2}{6} \tag{8.59}$$

The resulting signal-to-noise ratio between signal power and error power is

$$\text{SNR} = 10\log(3) - 20\log(\omega T_s) \tag{8.60}$$

An increase in sample rate or a decrease in sample time of a factor two results in 6 dB of signal-to-noise ratio or the equivalent of 1 bit. When discussing oversampling in Sect. 10.1, the increase in sample frequency in a converter must be a factor of four

higher to gain the same amount of signal-to-noise ratio. In a level-crossing device the time axis serves as the quantization axis and a frequency doubling doubles the accuracy. In the oversampled situation, the faster sampling results in a wider frequency band that only serves to spread out the quantization power.

The requirement that the signal passes no more than one quantization level in one sample period makes this principle less interesting for many application domains. A major implementation problem is that the delay between the actual level crossing and the resulting digital transition must be constant, which is not trivial as the delay of latches depends on the over-drive voltage, Sect. 8.1.3.

For low resolutions this principle converges towards asynchronous delta modulation, Sect. 10.8.2 [270].

8.9.2 *Feed-Back Asynchronous Conversion*

Asynchronous converters constantly monitor the signal and create in a feed-back path a best fitting replica. The converter in Sect. 8.9.1 can only change 1 level at a time. Some low-resolution high-speed examples are reported [271, 272] that allow a faster increase. An example is shown in Fig. 8.126. The non-clocked MSB comparator continuously monitors the signal. The MSB bit is fed into a simple digital-to-analog converter that supplies either a $1/4V_{ref}$ or $3/4V_{ref}$. The second comparator uses this input reference to determine the MSB-1. If a signal passes through its range, the digital code of the converter will follow. A settling time of a few nanoseconds for 6 bits suffices. This is an asynchronous successive approximation converter.

Comparable to the level-crossing analog-to-digital converter the interfacing to the (clocked) digital world introduces quantization errors.

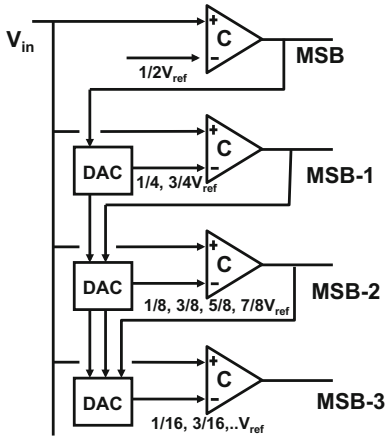


Fig. 8.126 The asynchronous successive approximation analog-to-digital converter [271]

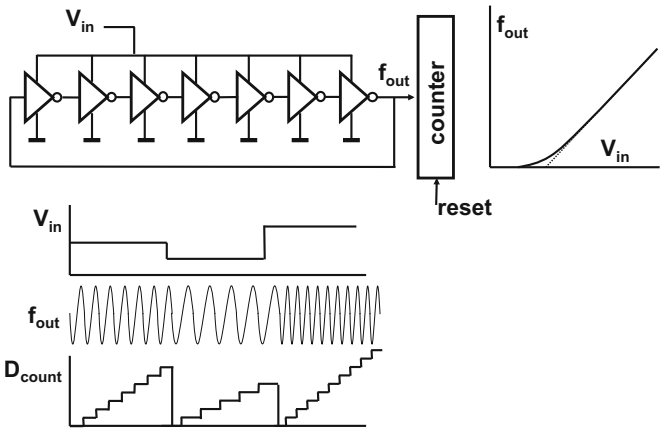


Fig. 8.127 VCO as voltage-to-frequency converter [273]

8.9.3 Time-Related Conversion

In some systems, where a full analog-to-digital converter is not the optimum solution for area or power reasons, the conversion of a voltage to an intermediate quantity (such as a frequency) can be a solution. A voltage controlled oscillator (VCO) is by principle a voltage-to-frequency converter. The frequency deviation is preferably in first order proportional to the voltage deviation. Both domains of a VCO, input and output, are time-continuous and amplitude continuous, see Fig. 8.127. However a frequency is easily mapped on the digital domain by using a time window to count the number of frequency periods. This is the quantization step. As the counter continues, the quantization error of the previous sample is added to the next. So the DC-content of the signal is preserved and this quantizer shows a first-order noise shaping, see also Example 10.2.

The resolution is proportional to the difference between the maximum and minimum number of pulses in the time window. An implementation problem is the reset pulse for the counter. If this pulse coincides with a VCO pulse, the result is corrupted.

The arrangement in Fig. 8.128 shows a pulse-based conversion: the conversion begins after a star pulse allows to count. Next to counting the cycles of the VCO, more resolution is obtained from latching the state of the VCO pulse at the moment the stop pulse is received. Figure 8.128 shows a time-to-digital converter with a coarse counter and a latched fine conversion. Naraghi et al. [264] reports 300 kHz bandwidth at almost 8 ENOB and an F.o.M. = 98 fJ/conv. level.

In other systems a pulse width contains the required signal information or a time interval must be monitored. This class of time-to-digital converters is based on principles resembling the counting analog-to-digital converter. A high frequency clock is counted during the interval of interest. In case this interval becomes too

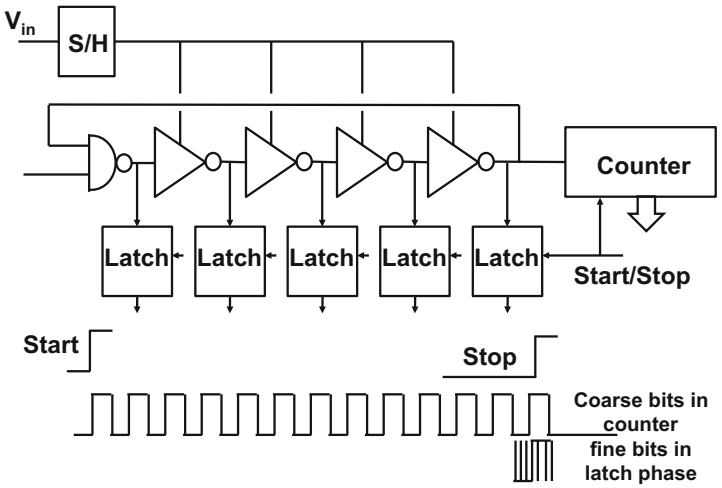


Fig. 8.128 Time-to-digital converter with coarse fine architecture [264]

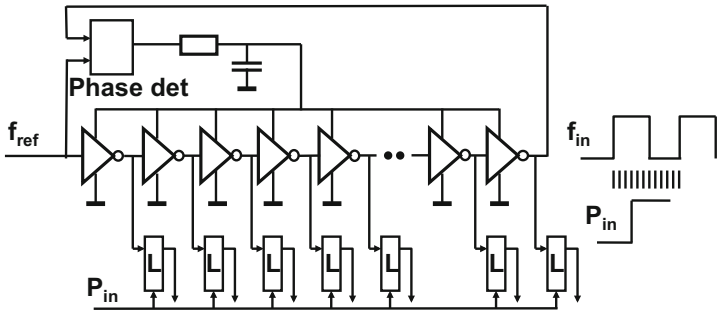


Fig. 8.129 Time-to-digital converter based on a delay-locked loop with a fixed reference frequency [274]

small (<100 ps), such a simple technique is not practical as the required counting frequency would become too high. Various structures as in Fig. 8.129 allow to subdivide the reference clock pulse f_{ref} and the resulting set of time-shifted pulses is used to clock the pulse P_{in} at a resolution of, e.g., $1/32$ of the clock period. This method of quantization is often applied in phase-locked loops. The accuracy is limited to the jitter in this system. Power supply variations and substrate interference can also influence the quality. An accuracy in the range of 3–10 ps_{rms} is possible [274].

An elegant implementation is the starving pulse converter [275]. The pulse that has to be measured is entered into a ring of inverters. If the inverters show perfectly symmetrical rising and falling edges without delay, the pulse would travel indefinitely in the inverter ring. One inverter is deliberately modified. Its rising edge is slow, so at each pass the pulse will become a time fraction shorter. The

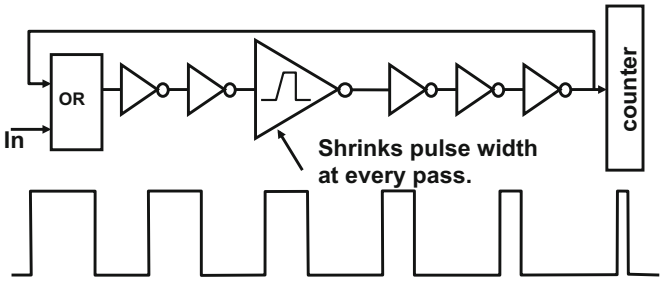


Fig. 8.130 Starving pulse time-to-digital converter [275]

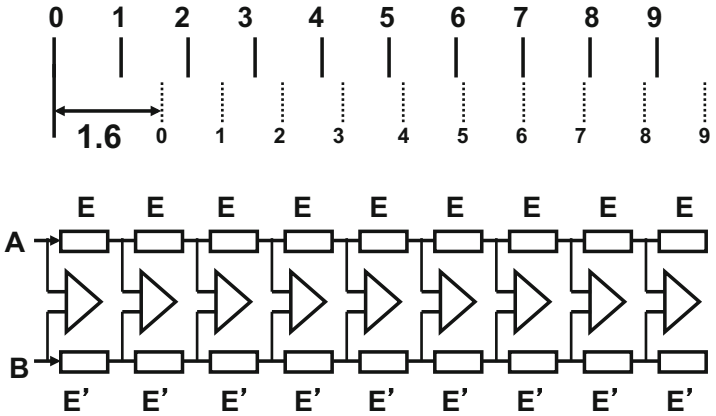


Fig. 8.131 The Vernier or Nonius principle

original pulse width is measured by counting the number the pulse before the signal is extinguished, see Fig. 8.130. The performance of this idea is perhaps inverse to its elegance.

8.9.4 Vernier/Nonius Principle

The need for higher resolution is not unique to the field of analog-to-digital conversion. In mechanics the Nonius or Vernier scale (named after sixteenth and seventeenth century Spanish and French mathematicians) is widely used to determine accurately the subdivision in a primary scale. Figure 8.131 (upper) shows a primary scale and a secondary scale (dashed). The secondary scale has a subdivision of 0–9 and its length spans 90 % of the primary scale. The unknown distance offsets the zero point of the primary scale from the zero point of the secondary scale. This unknown distance equals the entire number of primary scale units (1 in this example) plus that fraction which is denoted by the number in the

secondary scale where the marks of the primary and secondary scale are in line (In this example 6). This principle can be used in electronic designs as is shown in Fig. 8.131 (lower). The chains with elements E and E' form the primary and secondary scales while the comparators determine the position where both scales are "in-line." This principle can be used to convert a time period [276]. The elements are implemented as timing cells with slightly different delays. On terminals A and B the start and stop of the interval under measurement are applied. The same technique can also be used with elements E implemented as resistors forming a flash-like converter [277].

8.9.5 Floating-Point Converter

Most analog-to-digital converters operate on the assumption that over the entire range the same resolution is required. In some systems such as sensors and communication signals with a high crest factor,[20] the required resolution varies with the amplitude of the signal, see also the remarks on non-uniform quantization in Sect. 4.3.3 In computing, this sort of problems is addressed by means of floating-point arithmetic. Also in analog-to-digital conversion floating-point conversion is possible, although most system engineers prefer a fixed-point converter. Figure 8.132 shows an example of a floating point analog-to-digital converter. The actual analog-to-digital converter can use any architecture. The floating point mechanism around the converters consists in its basic form of an analog pre-scaler like a variable gain amplifier and a digital post-scaler. A processing unit detects whether the signal is sufficiently large to use the analog-to-digital converter optimally. If the signal is too large or too small, the processing unit will adapt the input and output scaling. If both these units run with inverse amplifications, the transfer curve will show a resolution-amplitude ratio that remains within certain limits. The difficult point in this design is the accuracy of the pre-scaler. The output

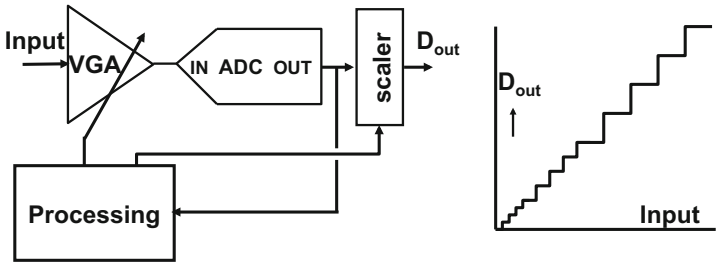


Fig. 8.132 The floating point principle realized with a variable gain amplifier (VGA)

[20] Crest factor is the ratio between highest amplitude and average amplitude.

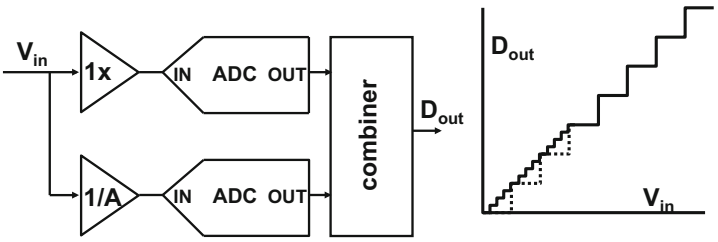


Fig. 8.133 The floating point principle implemented with two parallel converters

level with a gain of "2A" should be the exactly double of the gain "A." Offsets and gain errors are limiting this concept. A more detailed analysis is found in, e.g., [278]. The application of this sort of converters is in sensor interfaces.

The floating-point analog-to-digital converter in Fig. 8.133 uses two parallel converters. The upper converter acts on the small signal levels while an attenuated version of input signal is applied to the lower converter. The advantage with respect to the VGA solution is that no time is lost when a range switch must occur. Moreover, the logic in the combiner can easily adjust for undesired gain and offset errors between both signal paths, while making use of those signal levels that trigger both converters. Also forms of over-range can be applied.

Similar to other multiplexed circuits, the timing of the sample pulses must be accurately controlled as well as the matching of the bandwidths for large and small signals. Matching the $1\times$ buffer with the $1/A$ attenuator is a challenging task.

Exercises

8.1. In a 65 nm process, the input transistors of the comparators are designed with W/L $= 5/0.1\,\mu$m. The effective gate-drain capacitance is 0.2 fF/μm. Estimate the maximum ladder disturbance, if a 7-bit flash converter with a total ladder impedance of 1 kΩ is designed.

8.2. Deduce from Fig. 8.76 what is the maximum comparator offset that can be tolerated.

8.3. In a 6-bit flash 1 Gs/s analog-to-digital converter the wires of the clock lines must be kept as equal as possible. What is the maximum wireline difference that can be tolerated if the converter must operate at Nyquist frequency. Assume a propagation speed of 10^8 m/s.

8.4. Rank the following design parameters for achieving a high BER in order of importance: the gate width of the latch transistors, the length, the current, the gate oxide thickness, the gate-drain overlap capacitance?

8.5. Repeat the example of Table 8.2 with 0.25 μm technology data and an input range of 2 V. What is surprising?

8.6. Design a decoder for a 5-bit flash converter based on a Wallace-tree summation network.

8.7. Show that for a sinusoidal signal the digital output power of an N bit parallel output port is smaller than for a serialized single pin output.

8.8. At the input of an 8-bit analog-to-digital converter a ramp-type signal rising at 1 LSB per sample period is applied. What is the power consumption at the digital output compared to the maximum power. What happens to the power in case of 1 LSB DNL errors, and what in case of an 1 LSB INL error.

8.9. How can comparator mismatch in a multiplexed successive approximation converter affect the performance?

8.10. A 7-bit flash converter uses 128 resistors of 25 Ω each with a 50 fF parasitic capacitance to each comparator. If the internal signal swing of the comparator is 1 V, calculate the kick-back amplitude. Does the kick-back vary with input signal level? Where is the worst case level?

8.11. A sine wave must be converted with good absolute accuracy (without DC-offsets). No auto-zero or calibration mechanism is available. Give a reason why a flash converter is a better choice than a successive approximation converter.

8.12. In the Fig. 8.12 exchange the connections of the $\overline{clock}$ signal and the signals coming from the pre-latch. Does this solve the hysteresis problem and what is the cost?

8.13. Draw a transfer curve of a 2-bit coarse, 4 bit fine analog-to-digital converter. Add the transfer curve in case one of the coarse comparators has an offset of 1 % of the input range. Do the same if this offset applies to a fine comparator.

8.14. A 1.6 Gs/s analog-to-digital converter is built by multiplexing 64 successive approximation converters. If offsets are normally distributed calculate σ_{comp} for a 50 dB performance of 95 % of the dies. The bandwidth of the T&H circuits is 1.5 GHz. How much variation is allowed on this bandwidth?

8.15. A 1 LSB difference produces 10 mV voltage swing on the nodes of a latch in a comparator with parasitic capacitances of 100 fF. The maximum current in each branch of the latch is 100 μA. The current factor for a square transistor is 200 μA/V^2. Calculate the BER for a maximum decision time of 1 ns.

8.16. A comparator generates 50 fC of kick-back charge. The nodes of a ladder have a parasitic capacitance of 150 fF each. If the ladder can consume 1 mA from a 1 V reference source, and the converter must run at 2 Gs/s, what is the maximum resolution?

Chapter 9
Time-Interleaving

9.1 The Need for Time-Interleaving

In the previous chapter various conversion principles have been discussed. The fastest implementation of an analog-to-digital converter is the flash converter in Sect. 8.2, capable of reaching 4–6 Gs/s at ENOB $<$ 6 bit, Fig. 9.1. The penalty of this type of fast conversion is the lack of accuracy: static mismatch in comparators and ladders and dynamic errors in all components are hard to mitigate. Moreover every additional bit in resolution doubles the area, power, and input loading and requires additional current to avoid higher settling time constants in the ladders.

Improvement in accuracy without increasing the power is obtained by splitting the conversion process in various steps that are performed one after the other. Pipelining in Sect. 8.4 is applied to avoid too much speed loss. Nevertheless pipeline stages must do more work in one clock cycle than the simple decision taking of a flash converter. As subtraction of the reference value and a multiply step are required. The necessary settling time per pipeline stage is determined by the unity-gain bandwidth of the opamp, the required gain factor as set by the capacitor ratio and the number of settling time constants τ_{settle} needed to reach the desired accuracy (Fig. 9.2). More accuracy requires more time and reduces the sample rate and bandwidth. Gate delays, transmission line delay, and worst-case sign margins further limit the maximum speed. And as Fig. 9.1 shows, a gradual loss of speed with increasing resolution is the result. Section 8.4 shows several attempts to optimize this trade-off but pushing the sample rate above 1 Gs/s remains a challenge.

Yet, in advanced applications a higher speed/resolution performance is needed:

- Home cinema: consumers have bigger screens in their homes and want to operate the various channels in different windows. Some applications want to combine information from different channels. So in cable TV systems there is a need arising for full-spectrum cable TV (DOCSYS). A bandwidth starting at 48 MHz up to 1002 MHz must be converted at a resolution of 10 bit.

M. Pelgrom, *Analog-to-Digital Conversion*, DOI 10.1007/978-3-319-44971-5_9

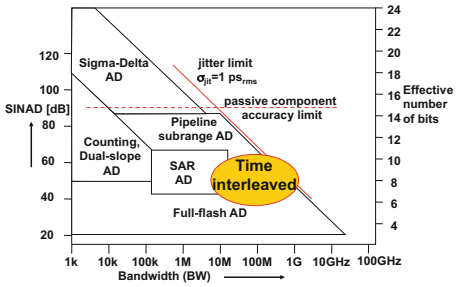


Fig. 9.1 This resolution-bandwidth graph shows the typical performance goal for time-interleaved converters

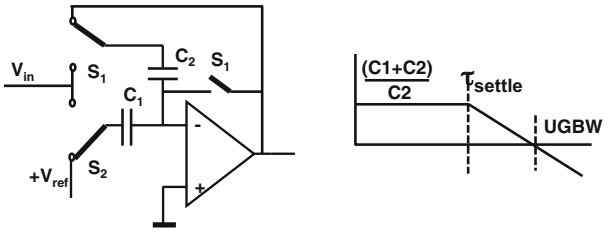


Fig. 9.2 A standard multiplying DAC section of a pipeline converter is speed limited by the feedback time constant at about half of the unity-gain frequency

- Direct RF sampling receivers for broadband applications replace the entire RF hardware section. The performance level is still low compared to the >100 dB sensitivity for radio broadcast. Applications in cable systems are being explored.
- Wireline communication: the speed of Ethernet over copper or fiber is going up rapidly, requiring: 10-25-100 Gb/s at an ENOB > 5 bit.
- Industrial: instrumentation and oscilloscope designers always want to process faster than the incoming signals. This tendency pushes the analog-to-digital converter for these applications to the forefront. Sample speeds are needed of 20–30 Gs/s at resolutions of 7–8 bit.

De-multiplexing or time-interleaving the signal over several identical analog-to-digital converters allows to match the high accuracy at limited speed of the basic structure to the wide bandwidth system requirements [279, 280]. Equivalent to the situation in digital circuits, demultiplexing will not reduce the number of steps to be taken, on the contrary the number of steps may increase. Keeping the power at reasonable levels is possible because the parallel processing can run at lower speed with lower (stand-by) currents and supply voltages.

9.1.1 Time-Domain Interleaving

In Fig. 9.3 a simple four-times interleaved analog-to-digital converter is shown. An analog-to-digital conversion section of a time-interleaved converter is often called a channel. The demultiplex factor N_{int} is here 4. The input multiplex switch changes

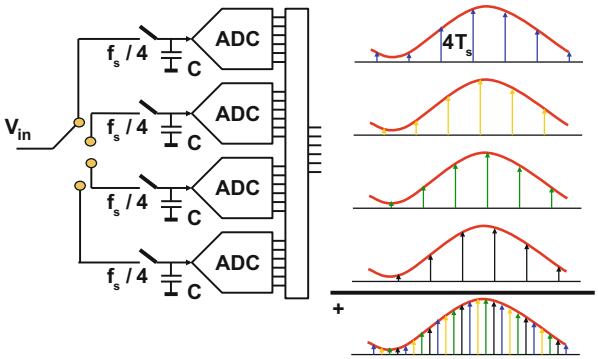


Fig. 9.3 A basic four-times interleaved converter in the time domain

position at the full sampling frequency f_s and each channel is connected to the input at a rate of f_s/N_{int}. As the right-hand side of Fig. 9.3 shows, the idea is that each channel receives the input signal but samples at an N_{int} fraction of the sample speed. Every channel samples at time-interleaved moments with respect to its neighboring channels. Despite the samples in the various channels are taken at different time moments, all four sample streams reconstruct the same signal with zero degree mutual phase shift. The overall result is found by combining the resulting digital samples in the correct order. The individual channel sampling violates the Nyquist criterion, but the overall conversion does not.

Time-interleaving can be applied to almost any type of analog-to-digital converter, be it that the advantages are not always the same. In [39] two pipeline converters are combined in a configuration that is sometimes referred to as "ping-pong," Kapusta et al. [281] uses two converters from a group of three SAR converters in order to scramble errors, in [282] eight flash converters are interleaved, in [283–285] 64 successive approximation converters are combined together. The choice for the number of interleaved converters, also called the interleave factor N_{int}, depends strongly on the type of basic converter and the speed that needs to be obtained.

Yet, a low or a high interleave factor often corresponds to different design philosophies. A low interleave factor $N_{int} = 2, 3, 4, \ldots$ often aims at obtaining a modest improvement in speed without sacrificing too much performance by re-using proven hardware blocks. The drawback of interleaving are the spurious tones caused by inequalities between the channels. These appear at and around integer fractions of the sample rate, e.g., f_s/N_{int}. For low interleaving factors such errors may be mitigated, calibrated, or even ignored.

The choice for a high N_{int} often comes from another angle. Some converter architectures show clear power efficiency advantages over others, but are unable to reach the high sample rates, e.g., SAR versus flash.The idea here is to combine many ($N_{int} = 16, 64, 512, \ldots$.) of those converters into a high-speed sampling analog-to-digital converter. In first order the overall sample rate f_s equals $N_{int}\times$ the sample

rate of the individual analog-to-digital converter $f_{s,i}$, so $f_s = N_{int} f_{s,i}$. If the associated track-and-hold provides a sufficiently large input bandwidth ($BW > f_s/2$), and the individual power efficiency is given by Eq. 4.25, the power efficiency of the overall conversion remains (largely) the same:

$$\begin{aligned} \text{Power efficiency overall} &= \frac{\text{Power}}{2^{ENOB} \times f_s} = \frac{\text{Power}/N_{int}}{2^{ENOB} \times f_s/N_{int}} \\ &= \text{Power efficiency per single ADC} \end{aligned} \tag{9.1}$$

This concept is illustrated at the hand of the F.o.M. plot in Fig. 9.4.

Power contributions of the overhead circuits have been ignored, but still this observation has a major consequence: the most power efficient conversion topology can now be chosen irrespective of the inherent latency. Provided that the input bandwidth is sufficient, its sampling speed can be increased by time-interleaving. Compared to structures with an inherently poorer power efficiency this allows to achieve high-speed conversion at relatively low power levels.

Obviously if a high N_{int} is chosen the consequence will be that the input buffer is heavier loaded. As all sample capacitors for each channel must meet the kT/C requirement, the driver must be able to deliver this power and maintain the signal quality level.

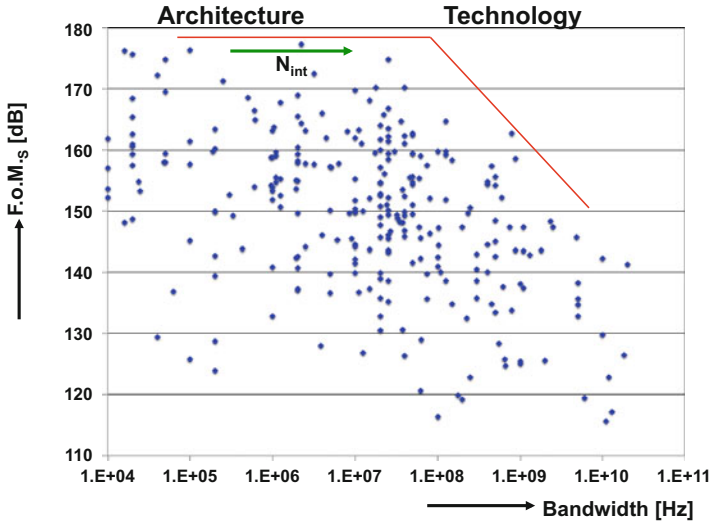


Fig. 9.4 Time-interleaving can be used to exploit the energy efficient, but slow architectures. In this plot of Schreier's Figure of Merit, high-efficiency converters at modest speed performance are located in the *upper left corner*. The interleaving approach (*the green arrow*) combines N_{int} of these to a new device with a higher bandwidth (Data from: B. Murmann, "ADC Performance Survey 1997–2015," Online: http://web.stanford.edu/~murmann/adcsurvey.html)

Example 9.1. In a ping-pong time-interleaved converter the interleave factor $N_{int} = 2$. An input signal $V_{in}(t) = V_A \sin(2\pi f_{in} t)$ with a frequency f_{in} close to $f_s/4$ is sampled by both channels. Show that the first alias components in both channels cancel.

Solution. Equation 2.7 gives the series formula for the sample sequence. The sampling process in the two channels with each a sample rate of $f_s/2$ can now be seen as a modulation of the input signal with the following terms:

$$1 + 2\cos(\pi f_s t) + 2\cos(2\pi f_s t) + 2\cos(3\pi f_s t) + \ldots\ldots$$
$$1 + 2\cos(\pi f_s t + \pi) + 2\cos(2\pi f_s t + 2\pi) + 2\cos(3\pi f_s t + 3\pi) + \ldots$$

Obviously, the phase shifts lead to alternating signs in the lower sequence, which implies that these terms will disappear. Still, the multiplication of the input with these sequences can be performed resulting in:

$$\begin{aligned} &V_A(\sin(2\pi f_{in} t) + \sin(2\pi(f_{in} - f_s/2)t) + \sin(2\pi(f_{in} + f_s/2)t) \\ &\quad + \sin(2\pi(f_{in} - f_s)t) + \sin(2\pi(f_{in} + f_s)t) + \ldots\ldots) \\ &V_A(\sin(2\pi f_{in} t) - \sin(2\pi(f_{in} - f_s/2)t) - \sin(2\pi(f_{in} + f_s/2)t) \\ &\quad + \sin(2\pi(f_{in} - f_s)t) + \sin(2\pi(f_{in} + f_s)t) + \ldots\ldots) \end{aligned}$$

illustrating the cancellation of the alias frequency at $f_s/2 - f_{in}$ and at $f_s/2 + f_{in}$.

Figure 9.5 shows the input signal and its first alias component. The input frequency remains the same, but the alias is modulated by the channel sample rate and takes the π radians phase shift of the modulating frequency, not the phase shift for a frequency $f_s/2 - f_{in}$ that is $T_2/2$ delayed.

9.1.2 Frequency Domain View

The simple picture of time-interleaving in Fig. 9.3 requires further extensions. How can time-interleaving be understood in the frequency domain? Figure 9.6 shows the spectra of the four different conversion channels. As the local sample rate per channel is f_s/N_{int} the channel spectra repeat at integer multiples of this sample rate. The depicted bandwidth in the plot is less than $f_s/8$ in order to avoid a messy picture, but this does not limit the overall validity of the argument. In full operation the signal bandwidth will be $BW < f_s/2$. A frequency domain amplitude plot shows only half of the information: the phase is missing. So the deliberate time difference in sampling moments between the channels must be taken into account by supplying the phase difference due to the sample moment difference. Equation 2.7 can be

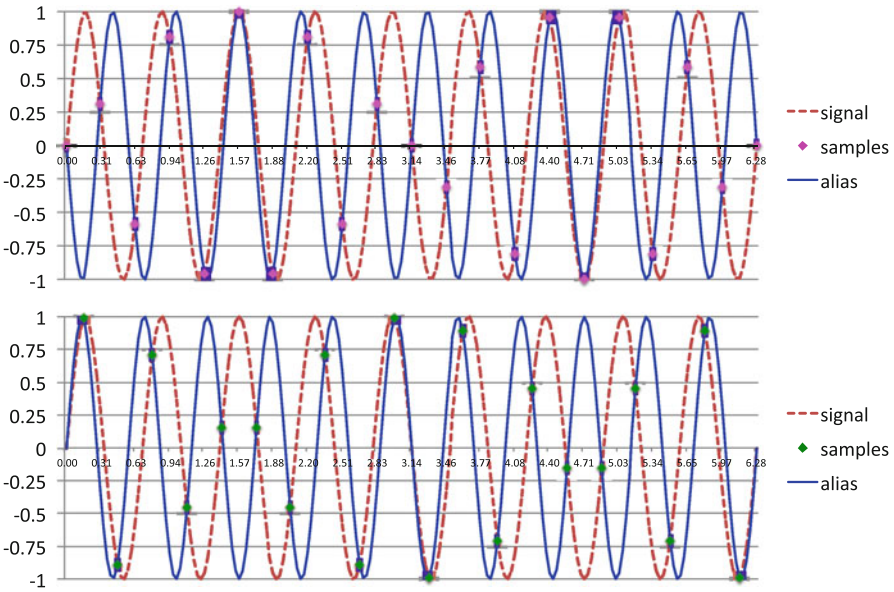


Fig. 9.5 The *dotted line* indicates the input signal to both channels. The sample points in both channels are interleaved and $T_s/2$ separated. The first aliases in both channels show a 180° or π radians phase shift

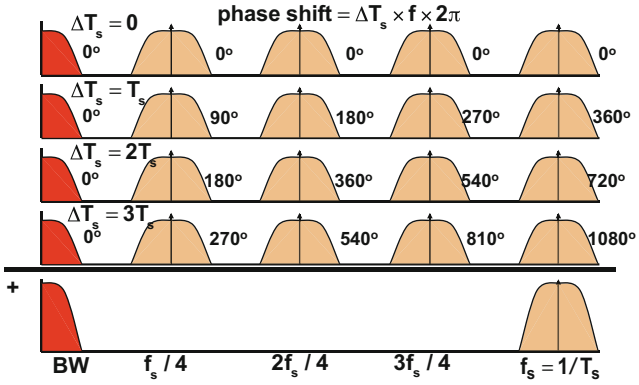


Fig. 9.6 A basic four-times interleaved converter consists of four channels. The spectra per channel are summed, taking their mutual phase relations into account

used to describe the sampling process per channel. Adding in this equation the time difference per channel $n_{chan}T_s$ allows to write the channel sample rate as a Dirac sequence and its Fourier expansion:

$$\sum_{n=-\infty}^{n=\infty} \delta(t-(nN_{int}+n_{chan})T_s) = \frac{1}{N_{int}T_s} \sum_{k=-\infty}^{\infty} e^{jk2\pi t/N_{int}T_s} e^{jk2\pi n_{chan}/N_{int}} \tag{9.2}$$

where $n_{chan} = 0, \ldots, (N_{int} - 1)$ is the channel number. The last term in the exponential equation represents the delay between the channels. In the frequency domain this results in a phase difference between channels and between integer multiples of the channel sample rate. Along the same lines as in Sect. 2.1 the resulting sampling spectrum for one channel is

$$\mathbf{A}(\omega, n_{chan}) = \sum_{k=-\infty}^{\infty} \mathbf{A}(f - kf_s/N_{int}) e^{jk2\pi n_{chan}/N_{int}} \tag{9.3}$$

The time-continuous spectrum $\mathbf{A}(f)$ is repeated at integer multiples of f_s/N_{int}, where each channel is shifted in phase by a fraction $k2\pi n_{chan}/N_{int}$ or a $k \times n_{chan}/N_{int}$ portion of 360°. In Fig. 9.6 the phase of the four channels in the example is given with respect to the first channel. A sampling moment difference of T_s will result at a frequency of $f_s/4$ ($k = 1, n_{chan} = 1, N_{int} = 4$) in a phase difference of $\pi/2$ or 90° of the second channel with respect to the first. In this simple example two sub-spectra with a mutual phase difference of π or 180° will annihilate. If this observation is applied to all combined sub-spectra around $f_s/4$, $2f_s/4$, and $3f_s/4$, the result will be zero. Only spectra around multiples of f_s are in-phase and will re-enforce each other and result in the desired sampling spectrum. The spectrum after perfect reassembly of the N_{int} channel spectra is

$$\mathbf{A}(\omega) = \sum_{k=-\infty}^{\infty} \mathbf{A}(2\pi(f - kf_s)) \tag{9.4}$$

as in Eq. 2.12. The next sections show how this break-up in channel spectra helps to understand the consequences of mutual inequalities.

9.2 Input Sampling

The crucial element in every time-interleave architecture is the distribution of the signal over the N_{int} channels. In this part of the design most of the offset, gain, and timing errors are encountered. Nearly all problems with sampling inaccuracy and bandwidth are also attributed to the input circuits.

9.2.1 Distribution

Distributing the samples over the different channels can be implemented in various ways. Figure 9.7 (left) shows a simple arrangement: a single track-and-hold function (e.g., a T&H circuit from Chap. 3) is employed that serves the various channels. This arrangement can be implemented in a manner that both the input sampling time

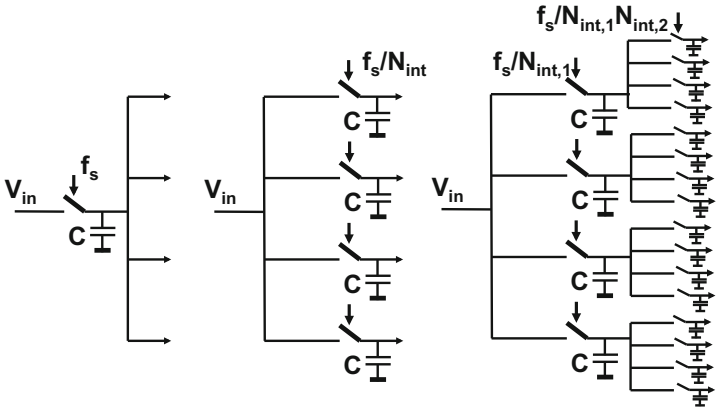


Fig. 9.7 Three different configurations for time-interleaving. *Left*: A single track-and-hold circuit serves the channels. *Middle*: Each channel uses its own track-and-hold. *Right*: Two stage time-interleaving

inaccuracy is limited to jitter problems and the bandwidth variation is mitigated. It should be noted that the parasitic capacitive load of the N_{int} succeeding stages is in parallel to the sampling capacitor. Often $1\times$ buffers are needed to mitigate coupling between channels. The basic requirements for the track-and-hold circuit do not differ from the simple analysis from Chap. 3. The error voltage equals:

$$V_{error} = V_{sample} e^{-T_{on}/R_{sw}C} \tag{9.5}$$

where R_{sw} represents the total resistive component in series with the switch and C the hold capacitor. For N-bit settling performance and 50 % duty-cycle of the sampling pulse, the T&H bandwidth must fulfill the requirements on the settling error:

$$e^{-T_s/2R_{sw}C} < 2^{-N} \quad \Rightarrow \frac{Nf_s}{BW} < \frac{\pi}{\ln(2)} = 4.5 \tag{9.6}$$

With $R_{sw} = 50\,\Omega$ and $C = 1$ pF the input bandwidth is $BW = 1/(2\pi R_{sw}C) = 3.2$ GHz. Now a simple trade-off exists: $Nf_s < 1.44 \times 10^{10}$ bits-Hz. A resolution of 10 bit limits the sampling rate to 1.44 Gs/s. At very high sampling rates this requirement on the switch may become hard to meet over all PVT[1] corners. However, for moderate bandwidths, this arrangement avoids timing inaccuracies or varying bandwidths due to the distributed hardware.

The small-signal bandwidth of the track-and-hold circuit poses a second constraint. This bandwidth should be sufficient in order not to attenuate the signal in the required input bandwidth specification.

[1]PVT: process, voltage, and temperature deviations from nominal process specification.

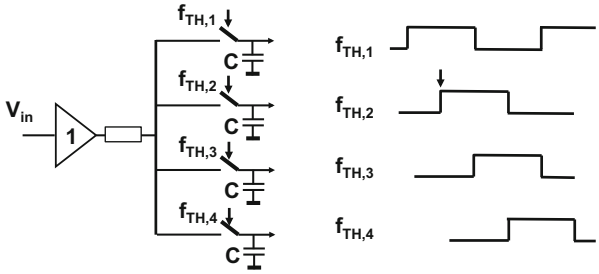


Fig. 9.8 A four-fold sampling scheme

$$BW = \frac{1}{2\pi R_{sw} C} > f_{in}$$

where f_{in} is the maximum frequency in the input bandwidth for which a specified attenuation is allowed, e.g., −1 dB.

The second arrangement in Figs. 9.7(middle) and 9.8 use a track-and-hold (T&H) circuit for each individual channel. Time-interleaving allows to stretch the sampling period. This relaxes the timing requirements of track-and-hold settling by approximately N_{int}. The thermal noise kT/C requirement applies to each individual T/H circuit, so the total capacitor area goes up by N_{int}. Figure 9.8 shows a four-fold sampling arrangement with stretched sampling periods. The settling time of the track-and-hold is now considerably longer and settling is easier. However, the small-signal bandwidth must now be met by a buffer loaded with two capacitors. Obviously more current is needed and the extra power is limiting the efficiency of time-interleaving.

Interference can occur as T&H2 switches when T&H1 is half way with its settling period. This arrangement loads the input buffer at any moment with two hold capacitors, causing a drop in input bandwidth [286, 287]. And a low bandwidth increases bandwidth mismatch sensitivity, see Eq. 9.19. With a distributed sampling clock and channel specific track-and hold circuit, these circuits will contribute to the issues discussed in Sect. 9.3: offsets, gain-errors, timing errors, and bandwidth errors.

The sampling process can be subdivided over multiple stages. A possible variant uses a two-stage track-and-hold structure as in Fig. 9.7 (right). This solution avoids to have many sampling capacitor circuits connected to one time-continuous input line, with all routing issues involved. The other argument for implementing a two-stage track-and-hold is the need to avoid sampling time errors, see Sect. 9.3.3 and bandwidth related errors in Sect. 9.3.4. With a proper timing of Fig. 9.7 (right) the first track-and-hold circuits will perform the time-continuous sampling while the second stage only sees a held-signal and is consequently insensitive to timing issues. Yet the idea to charge two capacitors in series seems not very attractive. When the two-stage architecture is applied with intermediate buffers [290] a good performance can be reached.

9.2.2 1× *Buffer*

In several designs the various issues with T&H circuits are addressed by inserting 1× buffers in the signal chain [39, 283, 284, 288–291]. These buffers allow lowering the loading of the hold capacitors and minimize any interference of the succeeding conversion process on the sampling. Moreover they provide decoupling between channels and avoid RF interference. In case of mutual interference between channels, the error correction and channel calibration turns into a multi-dimensional optimization problem. The obvious candidate for a 1× buffer is the standard source follower of Fig. 9.9. In a source follower configuration the transistor has its drain connected to the power supply. Now the source is connected to the load impedance and the circuit produces a close copy of the input signal on the output terminal. Power gain is achieved because the output current is larger than the input current. In a source follower the source is not tied to a fixed potential but carries signal. Therefore the equivalent transistor diagram contains also the source–substrate or source-back-bias transconductance g_{mb} and the output transconductance g_{ds}.

The input impedance is calculated as:

$$\begin{aligned}\frac{v_{in}}{i_{in}} &= \frac{1 + j\omega Z_L C_{GS} + g_m Z_L + g_{mb} Z_L + g_{ds} Z_L}{j\omega C_{GS}} \\ &= \frac{1 + Z_L g_m}{j\omega C_{GS}} + \frac{Z_L}{j\omega C_{GS}}(g_{mb} + g_{ds} + j\omega C_{GS}) \end{aligned} \tag{9.7}$$

For low frequencies the input capacitance equals $C_{GS}/(1 + g_m Z_L)$. The effect of the gate capacitor is reduced as the source largely follows the gate voltage change. So the net charge flowing into C_{GS} is low. If the load of the source follower Z_L contains a capacitance, the input impedance will show a strong increase at the frequency where the load capacitance reduces the $g_m Z_L$ term. The phase of the overall transfer will rapidly turn with 180°: 90° due to rapidly increasing input capacitance and another 90° in the load capacitor. These two close poles make a source follower a difficult element to handle in a feedback loop.

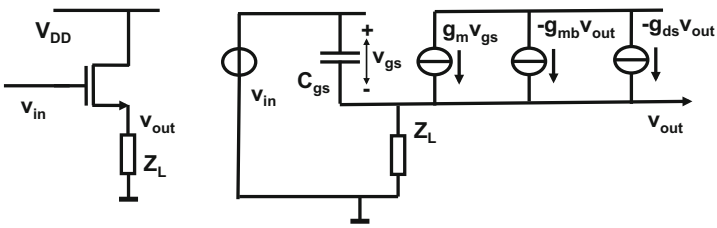


Fig. 9.9 The source follower circuit and its small-signal equivalent diagram

The output impedance on the source node is

$$\frac{v_{out}}{i_{out}} = \frac{Z_L}{1 + j\omega Z_L C_{GS} + g_m Z_L} \tag{9.8}$$

which goes to $1/g_m$ in practical design.

The transfer is found as:

$$\frac{v_{out}}{v_{in}} = \frac{Z_L(g_m + j\omega C_{GS})}{1 + j\omega Z_L C_{GS} + g_m Z_L + g_{mb} Z_L + g_{ds} Z_L} \approx \frac{g_m R_L}{1 + g_m R_L + g_{mb} R_L + g_{ds} R_L}\bigg|_{\omega=0} \tag{9.9}$$

Where R_L is the real part of Z_L. When a tail current source is used R_L is very high and:

$$\frac{v_{out}}{v_{in}} = \frac{1}{1 + g_{mb}/g_m + g_{ds}/g_m} \tag{9.10}$$

At low frequencies the transfer in a practical circuit is around 0.9.

The g_{mb} term in Eq. 9.10 can be eliminated when the substrate of the transistor is tied to the source, see Fig. 9.10 (left). This connection avoids the threshold voltage modulation due to the voltage dependent depletion layer and also the associated distortion. In standard processes, this connection limits the choice for source followers to PMOS devices as the local substrate of NMOS transistors cannot be separated from the bulk. More advanced technologies use "triple-well" constructions that allow NMOS transistors with their own isolated p-wells. In Fig. 9.10 (middle) the PMOS device implements the $1\times$ buffer [286]. The additional NMOS device acts as a secondary source follower keeping the drain–source voltage over the PMOS constant. This circuit eliminates the attenuation due to the output impedance g_{ds} in Eq. 9.10. Another solution uses a cascode of triple well devices [288].

The right most circuit in Fig. 9.10 is known as a flipped voltage follower [292]. Now the drain voltage of the follower M_1 is actively used in a feedback loop to the M_2 transistor. The gain approaches unity, but for high-speed operation the pole at the drain node of M_1 starts spoiling the performance.

None of these buffer circuits is free of problems.

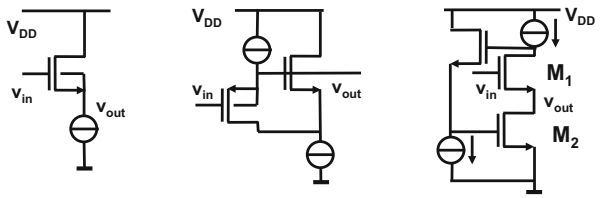


Fig. 9.10 Three implementations of a unity-gain amplifier

9.2.3 *Track-and-Hold Implementations*

The considerations for the T&H implementation are not very different from what has been discussed in Chap. 3. When SAR or pipeline converters are time-interleaved, the built-in T&H functionality is used. Yet, the very high frequencies and the advanced processes involved in time-interleaved design have resulted in some interesting realizations.

In [286] the problem of selecting the appropriate track-and-hold and not losing too much propagation time is solved by means of the circuit in Fig. 9.11. If the circuit is not selected, the bootstrap circuit is de-activated and the NMOS transistor driven by $f_{TH,i}$ will prevent any activity by the PMOS device. When the circuit goes in the track mode the bootstrap circuit is activated. The actual sample is stored on the falling edge on the sample transistor. This switch is activated by means of the differentially driven PMOS device.

The rather large gate-source capacitor creates in all source followers a feedback path from output back to input. The sample capacitor and buffer form therefore a second order system. During tracking the output follows the input. When the switch opens, a new charge equilibrium between the capacitors and non-linear parasitical capacitors is established. But as the output voltage lags the input, there will be a short settling period after the sample switch is opened. Limotyrakis et al. [39] analyses this problem in detail. Louwsma et al. [286] proposes a switch S disconnecting C_{out} at the sampling instant (Fig. 9.12).

More T&H circuits are found in Chap. 3.

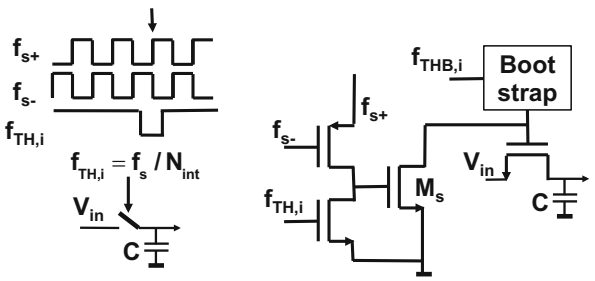


Fig. 9.11 An elegant implementation of the sampling switch [286]

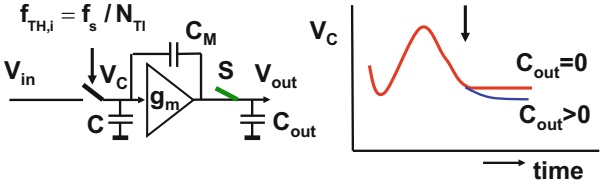


Fig. 9.12 A sampling circuit followed by a loaded source follower creates a second order response [39, 286]

9.3 Time-Interleaving Errors

Figure 9.13 displays time-interleaving of analog time-discrete circuits. A chain of analog-to-digital and digital-to-analog converters is shown. If all channels are perfectly equal, the above described cancellation of spectra will work. In reality the channels differ. The dominant disadvantage of time-interleaving is the unwanted output signal, consisting of tones and spectra due. The main issues are

- Different DC offsets between the channels will lead to an output pattern that repeats over a period $N_{int}T_s$.
- Gain differences between the channels will leave uncanceled portions of the sub-spectra.
- Sampling timing errors add up to the required time shifts and translate in phase errors between the sub-spectra.
- Bandwidth differences create also phase differences and limit the accuracy of cancellation.
- Timing differences during the reconstruction create also phase differences.

These unwanted additions to the output signal are caused by random and systematic errors. The problem in designing a chip with analog time-interleaving is that a two-dimensional lay-out has its limitations in achieving fully identical structures, e.g., a simple metallization crossings is inherently asymmetrical.

The systematic errors can be mitigated:

- Technological fixed errors are reduced if all multiplexed structures have the same orientation on the chip: no mirrored or rotated placements, see Table 5.5. Although it is important to keep identical circuits close together (e.g., in order to avoid gradients), a trade-off is necessary with respect to the proximity effects

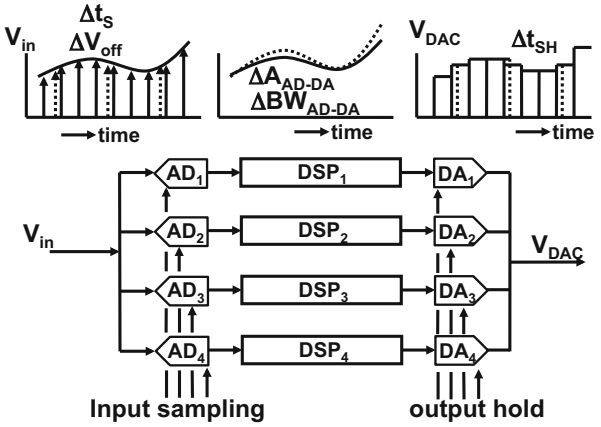


Fig. 9.13 The penalty of time-interleaving: inequalities between the channels

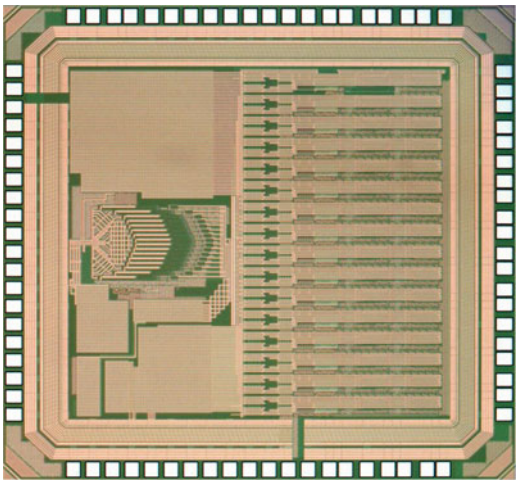


Fig. 9.14 On this time-interleaved chip the track-and-hold circuits (*left*) are positioned in a semi-circle to keep the distances equal [286]. The 16 channel pipeline-SAR analog-to-digital converters are on the *right* on this chip photo. Courtesy: S.M. Louwsma

of other structures. Many technological fixed errors cause either patterns at the multiplex frequencies or gain errors.

- Electrical errors such as voltage-drop errors are reduced by star-connected signal, pulse, and power wiring, see Fig. 8.29. Unequal power supplies lead to timing differences. RC time constants should be matched in value, but preferably consist of identical components. Digital and analog power supplies are separated, however, the common substrate is difficult to avoid and precautions in the digital part have to be taken as well.
- Timing errors affect in first order only the input sampling and the output restoration. The relative position of the pulses has to be accurate. Moreover, fast edges are important to avoid that the moment of sampling or holding becomes signal-dependent.

Distances have to be kept identical, see Fig. 9.14. The speed of signal propagation over ideal conductors surrounded by silicon and silicon dioxide SiO_2 is roughly $1/\sqrt{\epsilon_{SiO2}} \approx 2$ to $1/\sqrt{\epsilon_{Si}} \approx 3.5$ fraction of vacuum, so 7 ps corresponds to one millimeter. In case the wires show resistance, the speed even further reduces, and wire length differences must be brought back to a few micrometers.

In Sect. 7.6 techniques are discussed to avoid or mitigate additive errors. Here especially the multiplicative errors and timing errors are analyzed.

9.3.1 Random Offsets Between Channels

Random DC-offset in the N_{int} channels leads to a pattern in the output signal. There is some similarity with an array of parallel comparators as in flash converters, see Fig. 8.30. In flash converters these errors are linked to the comparator selected

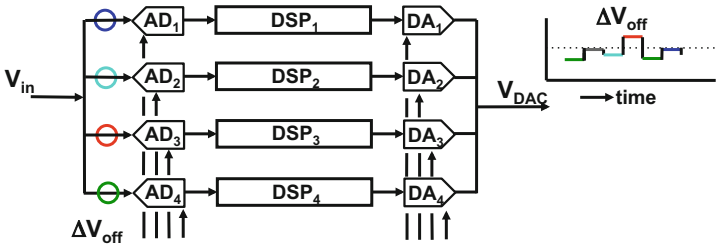


Fig. 9.15 Offsets in time-interleaved structures show up in the output as repetitive patterns

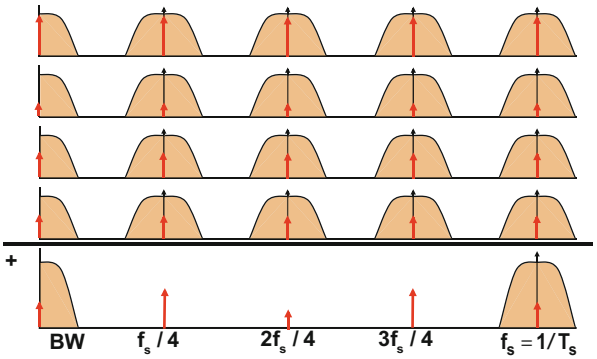


Fig. 9.16 Output spectrum of a four-times interleaved converter with offset errors generating discrete tones at multiples of f_s/N_{int}

by the signal amplitude, where in time-interleaved structures the random errors as addressed sequentially. Fixed timing sequences lead to fixed patterns, random selection of channels leads to (somewhat) random patterns.

Without precautions this leads to a repetitive error pattern in time, Fig. 9.15. If every offset can be modeled by one statistical process with variance σ^2_{Voff}, the overall power contribution to the signal is equal to the rms value of the error source σ_{Voff}. In the frequency domain this pattern is represented by fixed frequency components at multiples of the channel sampling rate f_s/N_{int}, Fig. 9.16. In the sampled data domain this results in:

$$V_{off}(t) = \sum_{n=0}^{\infty} A_{off,n} \sin(j2\pi n f_s t/N_{int}) \tag{9.11}$$

where $A_{off,n}$ is the random amplitude of the individual tones. Only the tones within one alias section (e.g., $0, \ldots, f_s/2$) have unique amplitudes, the tones in other sections have an identical amplitude pattern. After reconstruction in a set of digital-to-analog converters, this result is filtered with the applicable $\sin(x)/x$ function.

As the power in the time sequence must equal the expectation value of power in the frequency components:

$$\sigma_{Voff}^2 = E\left[\sum_{n=0}^{N_{int}-1} A_{off,n}^2/2\right] \quad \sigma_{Aoff}^2 \approx 2\sigma_{Voff}^2/N_{int} \tag{9.12}$$

Choosing a higher interleave factor reduces the individual tone amplitude, but the overall power remains the same. In a structure with a sufficiently large number of channels random channel selection will turn the tones into a sort of white noise, whose total power within $0, \ldots, f_s/2$ equals the error source power.

Image sensors present an example of extreme interleaving, see Fig. 9.17. Millions of photo sensitive cells form the imaging array, consisting, e.g., of 4000 columns with each 3000 cells. Column offset errors turn into stripes in the image and create "fixed-pattern noise." The eye is sensitive to static fixed error amplitudes of over 0.1 % full-scale. The human eye also provides an escape: random errors are only visible if the magnitude is above 0.5 % full-scale. So various randomizing swapping tricks are applied to reduce the visibility of these offsets.

Randomizing DC-offsets can reduce the amplitude of fixed tones; however, the error power remains in the conversion system and ultimately sets limits to the performance. Removing errors is (when possible) a better strategy. Some designs calculate the mutual DC-differences and subtract this error in the digital domain. In other designs a feedback is used.

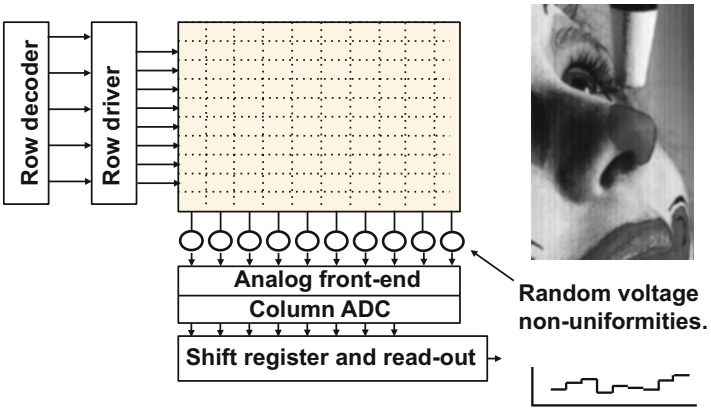


Fig. 9.17 Image sensors are extreme interleaved structures. Here offsets show up as stripes in the picture [265]

9.3.2 Random Gain Differences Between Channels

Gain differences between channels consisting of analog-to-digital converters can have various origins. SAR and pipeline converters use capacitive ratios for their operation. In some topologies differences in these ratios result in gain errors. However, the most likely cause for a gain difference is due to errors in the reference supply, the biasing in the intermediate amplifiers, and buffer stages in the preceding demultiplexing. In many cases there is no perfect solution to remove these inequalities and the performance depends on balancing the (dis)advantages.

An example of the dilemma a designer faces is illustrated in Fig. 9.18. In the simple upper scheme, the distribution of currents for, e.g., an amplification stage is achieved via a mirroring configuration. As shown in Eq. 5.24 the detrimental effects of mismatch (noise and $1/f$ noise) can be reduced by choosing a large transistor length. However, the capacitive coupling of the internal voltages to the bias line may easily upset the bias voltage and couple signals from one channel into another. The lower distribution network uses an intermediate stage, adding additional mismatch and noise sources but achieving good decoupling between the channels.

An amplification error ΔG_k in line k leads to some uncanceled spectrum (Fig. 9.19):

$$\mathbf{A}_{\Delta G}(f, n_{chan}) = \frac{\Delta G_k}{N_{int} G} \sum_{k=-\infty}^{\infty} \mathbf{A}(f - k f_s/N_{int}) e^{jk2\pi n_{chan}/N_{int}} \tag{9.13}$$

Figure 9.20 shows the result.

The spurious tones in the frequency band $0, \ldots, f_s/2$ behave similarly as the alias components in a standard sampling process. Figure 9.21 zooms in on the spurious components around the two lowest multiples of f_s/N_{int} for an input signal $V_{in}(t) = V_A \sin(2\pi f_{in} t)$. The magnitude of these signals is attenuated because the contribution of the residue spectrum appears only during once every N_{int} samples at the output [280, 293].

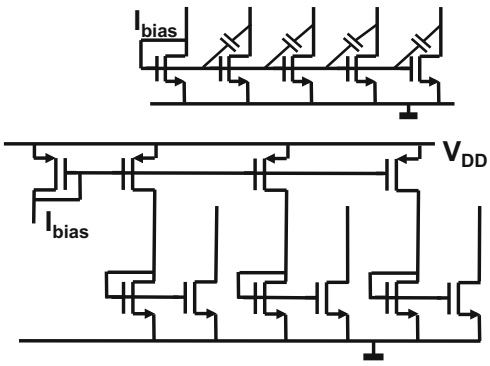


Fig. 9.18 Two methods to distribute a current

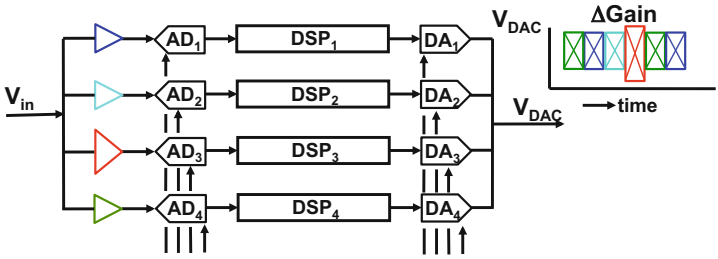


Fig. 9.19 The gain in the third channel differs, causing spurious tones

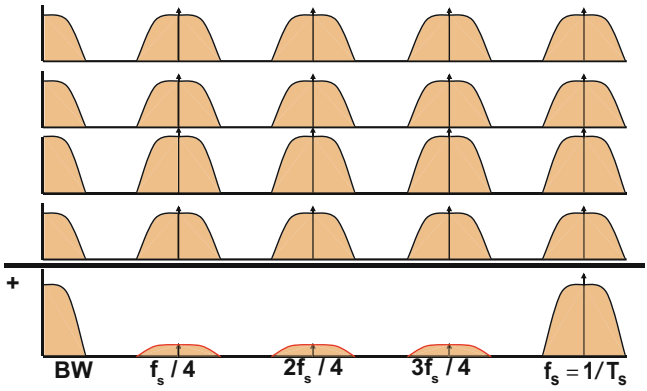


Fig. 9.20 The gain in the third channels differs, causing spurious spectra

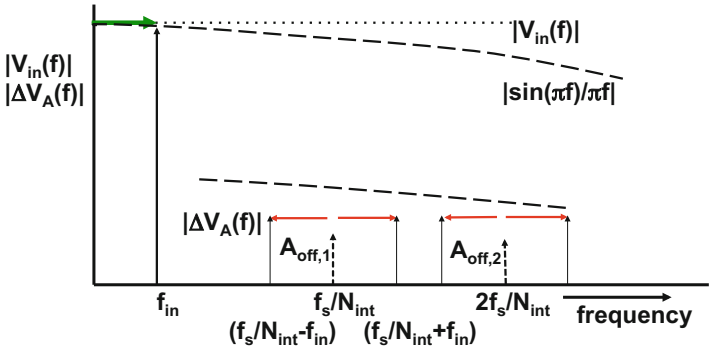


Fig. 9.21 The trajectory of the spurious components (*red arrows*) in the sampled data domain due to a gain error on the input signal (*green arrow*). After reconstruction the $sin(x)/x$ attenuation of the digital-to-analog converter differs for the signal and the spurious tone

In interleaved conversion systems the gain errors can be estimated from the channel power spectrum. Correction in either the digital domain or via an analog feedback path is possible.

9.3.3 Input Sampling Errors

Static timing errors between the input sampling devices of the channels result in a sampling sequence containing an error term. If channel number $n_{chan} = n_{\Delta Ts}$ has an extra delay of ΔT_s, then its channel sampling sequence is

$$\sum_{n=-\infty}^{n=\infty} \delta(t - (nN_{int} + n_{\Delta Ts})T_s - \Delta T_s) \tag{9.14}$$

In the sampled data spectrum this delay causes phase shifts for the copy bands around the multiples of f_s/N_{int} (Fig. 9.22):

$$\mathbf{A}(\omega, n_{\Delta T_s}) = \sum_{k=-\infty}^{\infty} \mathbf{A}(f - kf_s/N_{int}) e^{jk2\pi n_{\Delta T_s}/N_{int}} e^{jk2\pi \Delta T_s/T_s} \tag{9.15}$$

Small phase shifts on themselves are generating only minor errors under the condition that the sample is processed and restored with the same delay/phase-shift. The error appears because the delayed value is used as if it is valid for the time moments specified by the original ideal sampling sequence $\delta(t - (nN_{int} + n_{\Delta Ts})T_s)$. If an input sine wave $V_{in}(t) = V_A \sin(2\pi f_{in} t)$ is sampled at shifted time moments, an amplitude difference of:

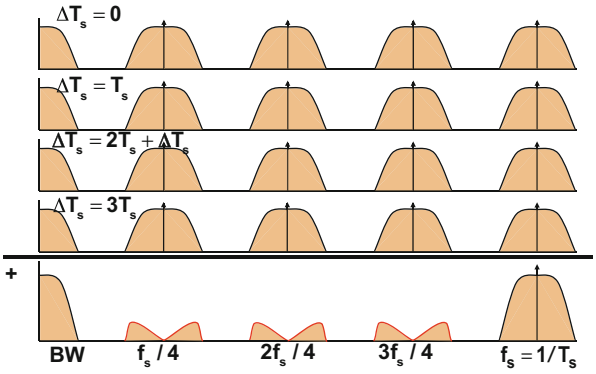


Fig. 9.22 A static timing error leads to errors in the cancellation of the spectra

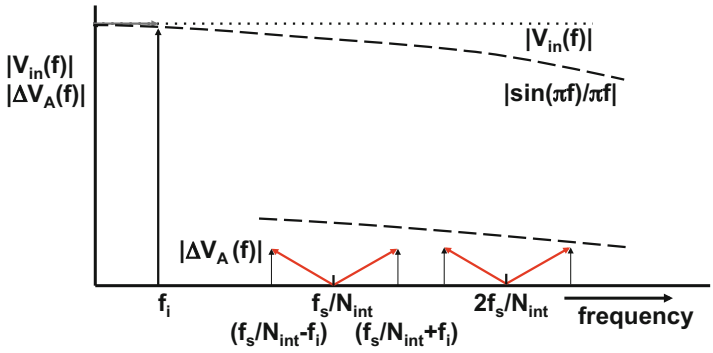


Fig. 9.23 In a time-interleaved converter with a input sampling error an input sine wave will generate spurious components that are proportional to $f_{in}\Delta T_s$. The trajectory of these spurious components starts at the multiples of the channel sampling frequency and grows with f_{in}. After reconstruction both amplitudes are attenuated with the $sin(x)/(x)$ function for their respective frequencies

$$\Delta V_A(t) = 2\pi f_{in} V_A \Delta T_s \cos(2\pi f_{in} t) \tag{9.16}$$

will occur in a similar fashion to the jitter analysis. This input sine wave at frequency f_{in} will generate tone at frequencies $kf_s/N_{int} \pm f_{in}$ and a power ratio

$$\frac{P_{\Delta V_A}}{P_{in}} = \left(\frac{2\pi f_{in}\Delta T_s}{N_{int}}\right)^2 \tag{9.17}$$

This power of this error is input-frequency dependent, and the spurious component grow with increasing input frequency, see Fig. 9.23. If measured after reconstruction, an additional term must account for the difference both the signal and the error signal experience from the zero order hold function.

Single timing errors that appear for low interleave factors N_{int} generate error patterns that are strongly related with the input signal. For large N_{int} values, the errors due to the individual timing errors smear out. These errors behave as a "random skew" and can be approximated by a random timing distribution with a variance $\sigma^2_{\Delta Ts}$. The relation of between this variance and the achievable signal-to-noise ratio, see Fig. 9.24 in such a time-interleaved architecture, resembles the behavior for jitter. The main difference is that jitter is a time varying phenomenon, while this "random skew" is fixed and can lead to reproducible errors for certain signals in combination with certain switching sequences.

Random skew applied to wide-band spectra will show a somewhat better performance than for a single tone at the highest frequency in that band. Dalt et al. [27] predicts for time-jitter an improvement up to a power factor of 3.

Example 9.2. A metal line runs over a chip. How much time delay is associated with a length $L = 100\,\mu\text{m}$?

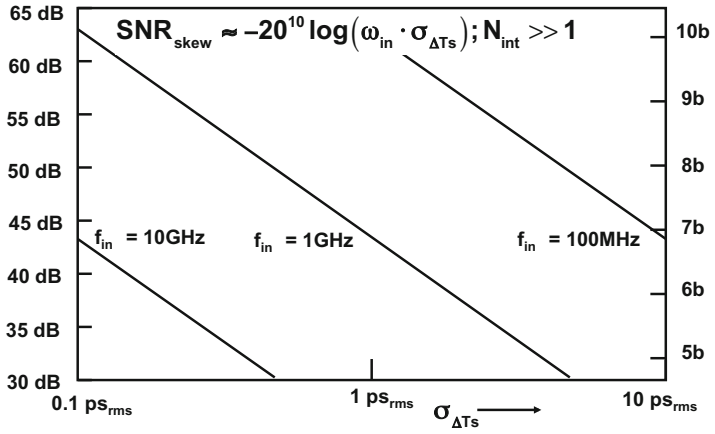


Fig. 9.24 For large N_{int} values, the individual timing errors can be approximated by a distribution with a variance $\sigma^2_{\Delta Ts}$. The relation between the variance and the achievable signal-to-noise ratio resembles the relation for jitter

Solution. Two aspects of the wire must be taken into account. First the wire transports an electro-magnetic wave. The inductance per unit length and the capacitance per unit length model the properties. From theory the propagation speed is

$$v_{wire} = \frac{1}{\sqrt{\epsilon_{SiO_2}\epsilon_0 \mu_{SiO_2}\mu_0}} = \frac{v_{vacuum}}{\sqrt{\epsilon_{SiO_2}}}$$

where $\epsilon_{SiO_2}\epsilon_0$ stands for the relative permittivity in SiO_2 and the absolute permittivity in vacuum. $\mu_{SiO_2}\mu_0$ correspond to the magnetic permeability and $v_{vacuum} = 3.10^8$ m/s is the velocity of light in vacuum. Silicon dioxide has the magnetic properties of vacuum and a relative permittivity coefficient of 3.9. The electromagnetic waves travel at $v_{wire} \approx v_{vacuum}/2$ over a wire. The time needed equals distance divided by speed and ends up for 100 µm at 0.7 ps. The second aspect concerns the resistivity of the wire surrounded by silicon dioxide. With a specific capacitance of $c = 1$ fF/µm and a resistivity of $r = 1\,\Omega/\mu$m, the heat equation 7.19 gives as a time constant for an open ended wire:

$$\tau = \frac{4rcL^2}{\pi^2}$$

Now the delay is 4 ps. Note that the wave propagation delay is proportional to the length, while the resistivity dominated delay goes with the square of the length. At this length the resistivity dominates.

Example 9.3. The sample pulse for each channel is buffered with separate circuits like in Fig. 9.25. The transition time on all nodes is 50 ps. If one buffer is fed with 60 mV less power supply than the other buffers, what is the resulting delay difference.

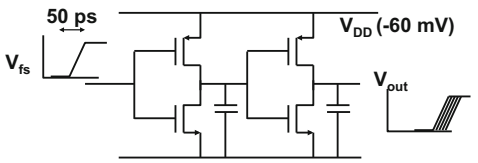


Fig. 9.25 What is the delay difference if the power supply of 1.2 V of this buffer drops by 60 mV

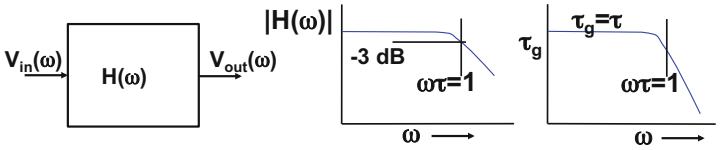


Fig. 9.26 A first order transfer has an attenuation of −3 dB when the radial frequency equals $1/\tau$. From DC till this frequency the group delay is almost constant and equal to τ

Solution. If half of the power supply voltage is used as the point where to measure the delay, each stage generates 25 ps. With 60 mV less drive voltage, the transition time will go up proportionally as will the delay. So the additional delay per stage is 1.25 ps, or 2.5 ps for both.

9.3.4 Bandwidth Differences

An additional class of errors between channels are deviations in frequency dependent signal processing, like the demultiplex network. The most pronounced difference occurs between the bandwidths of the sections that constitute a multiplexed structure. These differences will also give rise to input frequency dependent errors.

The group delay of a first order system is specified in Fig. 9.26 and by the standard equations:

$$H(\omega) = \frac{1}{1 + j\omega\tau} \qquad |H(\omega)| = \frac{1}{\sqrt{1 + \omega^2\tau^2}} \qquad \Phi(\omega) = -\arctan(\omega\tau)$$

$$\text{Group delay: } \tau_g(\omega) = -\frac{d\Phi(\omega)}{d\omega} = \frac{\tau}{1 + \omega^2\tau^2} \tag{9.18}$$

For signals below $\omega = 1/\tau$ a first order roll-off transfer behaves as a constant delay τ.

Obviously the bandwidth errors require the analog signal to be time-continuous. In an time-interleaved analog-to-digital converter, therefore the input track-and-hold circuits used for the demultiplex operation are the most sensitive parts for bandwidth variations. Track-and-hold circuits with different bandwidths show different group delays and the same analog signal will experience unequal delay times when passing

through these networks. If the input consists of a sine wave $V_{in}(t) = V_A \sin(\omega_{in} t)$, the error signal between two channels n and m can be approximated by an additional term:

$$\Delta V_A(t) = \frac{\partial V_{in}(t)}{\partial t}(\tau_{g,n} - \tau_{g,m}) = V_A \omega_{in}(\tau_{g,n} - \tau_{g,m}) \cos(\omega_{in} t) = V_A \omega_{in}(\tau_n - \tau_m) \cos(\omega_{in} t)$$

The resulting error is proportional to the input frequency and bears mathematically some similarity with the analysis of random sampling skew, see Eq. 9.16. For large N_{int} this error behaves again as a random variable. If the variation per channel is characterized as $\sigma_{\tau g}$, the attainable signal-to-noise ratio is limited to:

$$\mathrm{SNR} = \frac{\mathrm{f_{in}}}{\mathrm{BW}} \frac{\sigma_{\tau g}}{\tau_g} \tag{9.19}$$

Obviously the bandwidth variation must be minimized, see Fig. 9.27

Example 9.4. The bandwidth of a 64-times interleaved T&H circuit is nominally equal to the sample rate and determined by the on resistance of the switch and the hold capacitor. The first component shows a relative variation of 0.75 % while the second spreads with 1 %. What resolution can be expected from the uncalibrated set-up?

Solution. The group delay is given by Eq. 9.18 and approximately equal to $\tau_g \approx RC$. Applying the procedure given by Eq. 5.11, the result is

$$\frac{\sigma_{\tau g}^2}{\tau_g^2} = \frac{\sigma_R^2}{R^2} + \frac{\sigma_C^2}{C^2} = 0.0125^2$$

This will allow a signal-to-noise ratio of 1/2 (from the frequency/bandwidth ratio) times $\sigma_{\tau g}/\tau_g$, resulting in 1/160 or 44 dB.

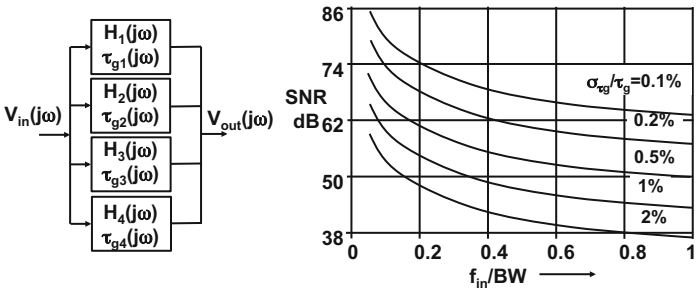


Fig. 9.27 Maximum signal-to-noise ratio as a function of input frequency and relative variation per channel, after [286]

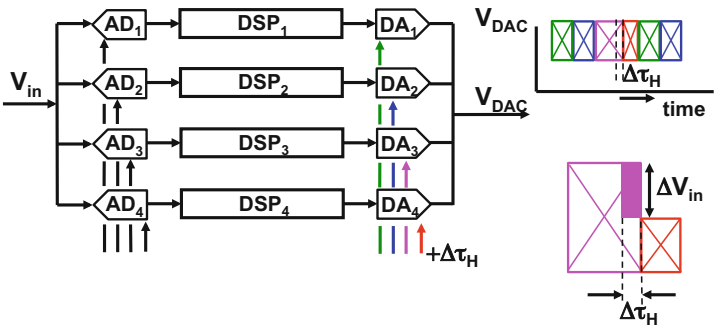


Fig. 9.28 The hold periods in the reconstruction are not equal

9.3.5 *Reconstruction Errors*

Figure 9.28 shows a time-interleaved structure with one reconstruction clock shifted by $\Delta\tau_H$. One sample will be held too long while the subsequent sample is too short. A simple approximation approximates the error as the maximum difference between two sample processed by a zero order hold function with hold time $\Delta\tau_H$ repeated at a rate of f_s/N_{int}.

The maximum difference between two samples is

$$\Delta V_{in}(t) = \frac{\partial V_{in}(t)}{\partial t} T_s = V_A \omega_{in} T_s \cos(\omega_{in} t)$$

and the transfer function of the zero order hold is

$$H_{z-o}(f) = \frac{\sin(\pi f \Delta\tau_H)}{\pi f T_s} \approx \frac{\Delta\tau_H}{T_s}$$

and the power ratio of the error signal versus the desired signal:

$$\frac{P_{\Delta\tau_H}}{P_{in}} = \left(\frac{2\pi f_{in} \Delta\tau_H}{N_{int}}\right)^2 \tag{9.20}$$

In case the reconstruction clock is jittering and all switching moments are subject to a deviation from a distribution with variance $\sigma^2_{\tau_H}$ the SNR formula moves (in a first order approximation) to the same description as for jitter: SNR$=$ $20\,^{10}\log(2\pi f_{in}\sigma_{\tau_H})$.

In [294, 295] a two-time-interleaved digital-to-analog converter is presented running at 11 Gs/s. Here the interleaving errors are of the same magnitude as the distortion.

9.4 Time-Interleaving Architectures

In a first order analysis, time-interleaving allows to push the conversion speed up to the technology limits, while the power efficiency is maintained. The main speed limitation is given by the sampling process and the signal distribution as here the full bandwidth and signal quality must be reached. After sampling there will be more time per decision allowing to optimize power, relax the design margins, and other sampling time dependent parameters in an analog-to-digital converter such as BER.

A second look: multiple T/H capacitors and overlapping timing can lead to higher input load and drive power. More hardware is more capacitance. Distribution, synchronization, and combining of digital signals cost power. And then the signal and timing matching issues need to be addressed.

The most popular converters for time-interleaved architectures are the pipeline and successive approximation topologies. Their inherent sampling mechanism fits well to the time-interleaving needs. Pipeline converters can push the speed performance up, while the successive approximation converters exploit their power efficiency.

Some applications have their own boundary conditions. In image sensors, many thousands of converters are used in parallel to process a position-interleaved signal. For this application the linear or counting converter is an often chosen solution.

Still many trade-offs are possible as will be discussed in the next sections.

9.4.1 Ping-Pong Architecture

The most simple form of time-interleaving uses two channels $N_{int} = 2$ and is often nicknamed ping-pong architecture, e.g., [39, 179, 296–298]. The architecture in Fig. 9.29 uses the even and odd pulses of the overall sample clock. Time-interleaving requires here a proven analog-to-digital converter design in combination with some clock circuitry and a digital multiplexer. The errors associated with

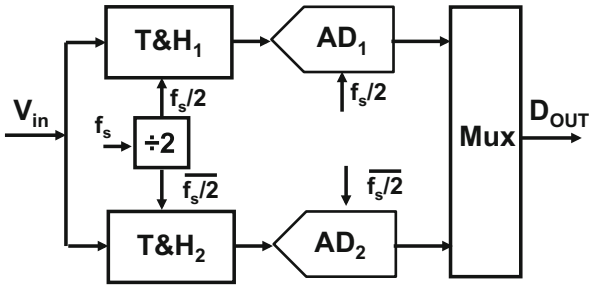


Fig. 9.29 The sample rate is divided by two and in opposite phase the samples are taken and processed at half-rate speed

time-interleaving are easily identifiable. Offset between the two channels results in a $f_s/2$ component. In some systems this spur can be ignored as it will not interfere with in-band frequencies. Gain and timing errors are very localized and can be adjusted with some analog trimming or digital correction.

Obviously the edges of the two half-rate local sample clocks that determine the sample moment in Fig. 9.29 must be perfectly timed. One solution uses the overall sample frequency and divides this frequency by two. Other solutions use directly a half-rate frequency signal and try to delay one edge to obtain a perfect 50 % duty cycle, as a primitive PLL.

For high-speed performance often two pipeline converters are employed in ping-pong mode [296]. Pipeline and successive approximation converters have built-in T&H circuits, reducing vastly the complexity of the interleaving. Note that each individual sample on a capacitor experiences kT/C noise. Each sample structure on itself must therefore fulfill the SNR specification.

Pipeline converters are constructed of multiplying DAC sections (MDAC) as described in Sect. 8.5. The MDAC variant based on opamp sharing samples in one clock phase by means of a passive network, while the amplification and subtraction is performed in the second phase. This opamp-sharing concept of Fig. 8.79 (right) is modified to Fig. 9.30 and fits nicely to the ping-pong architecture as a single opamp and two capacitor networks can service the two channels. A telescopic opamp is a candidate to implement the high-gain high bandwidth amplifier [223]. Settling errors show up as inter-symbol interference and are corrected with digital filters. The design uses offset, gain, and timing correction and performs at a 9 effective bit level at 5.4 Gs/s [296].

A popular implementation of the pipeline architecture employs successive approximation converters, see Fig. 9.31 [179, 286, 297]. Each channel consists of two clocked pipelined successive approximation sections connected by an amplifier. In [286] the amplification factor is 16 and with two 6-bit converters a large overrange is created that is spent on allowing slow settling in the SAR. With 0.4 ps_{rms} sampling jitter at an input frequency of 3.6 MHz the resolution is 6.5 ENOB.

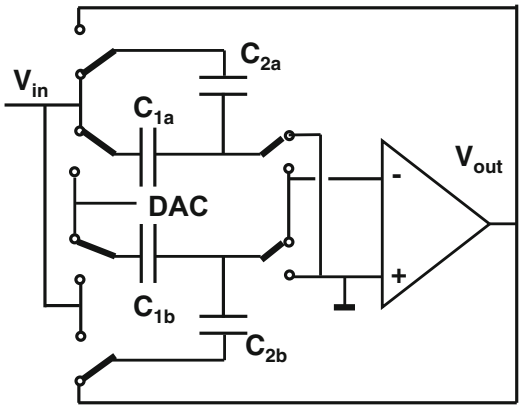


Fig. 9.30 An extended multiplying DAC (MDAC) structure is based on Fig. 8.79 (right) and fits well to the ping-pong architecture [296]. The upper capacitors C_{1a}, C_{1b} track the signal while the lower capacitors C_{2a}, C_{2b} are switched in amplification and subtraction mode. In the following clock phase the roles are alternated. The opamp has a very high UGBW to avoid inter-symbol interference

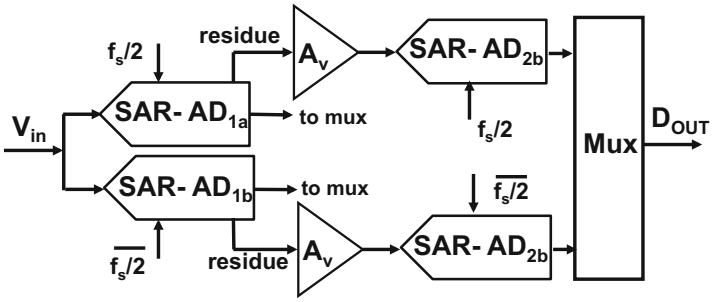


Fig. 9.31 Two SAR converters are pipelined in a channel for optimum conversion efficiency [179, 297]

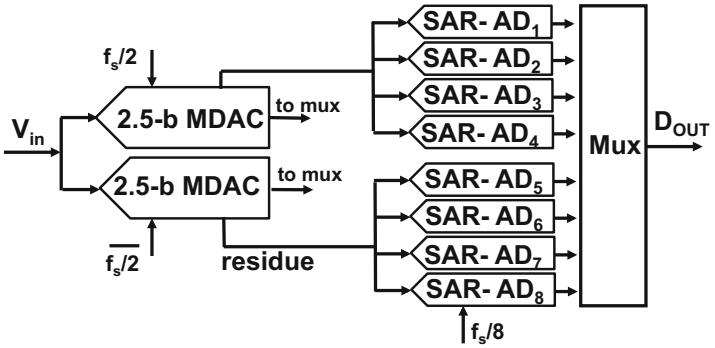


Fig. 9.32 Two MDAC front-end stages operate in ping-pong mode, their residues are fed into 8 SAR converters [298]

In [297] the two channels construct a fully dynamic ping-pong architecture. Each channel is build with two self-timed pipelined SARs. One channel is in a sampling phase while the other performs a self-timed conversion cycle. An intermediate 4× amplifier is used to separate both SARs and ease the requirements on the second converter. As the second converter applies some overrange, any misalignment between both SAR sections can be calibrated out. The performance is limited to 250 Ms/s at 9 bit effective bits resolution with an F.o.M of around 10 fJ/conv step [297].

The same concept is the basis for [179]. The amplifier connecting the two SAR converters into a pipeline is here based on charge integration. At 80 Ms/s the SINAD for Nyquist frequencies is 66 dB (10.7 ENOB), mainly limited by interleave tones, attributed to gain mismatch [179]. The F.o.M. is comparable to the SAR approach in [297].

Successive approximation converters allow low-power operation but tend to be rather slow. The ping-pong architecture improves the speed by a factor of two. In the architecture [298] of Fig. 9.32 a fast pipeline MDAC section with six quantization levels is operated with a shared opamp topology as in Fig. 9.30 [296]. The amplified

residue is fed into a bank of eight successive approximation converters, running at $f_s/8$ and exploiting their low-power property. The 8-bit SAR relies on the inherent matching properties of the capacitors and needs no error correction. The SAR range is calibrated to fit to the MDAC. More calibration is needed as the sample-and-hold less MDAC requires timing adjustment of the quantizer to the signal sampling. For power optimization the MDACs are fed from a 1.8 V supply while the SAR runs at 1 V supply. The overall sample rate is now 5 Gs/s at 8.3 effective bits and 150 mW power in a 28-nm CMOS process [298].

9.4.2 Low N_{int}

The ping-pong architecture realizes a two-fold speed improvement at an effective resolution of 8–10 bit. The main limitation is the inequality between the two channels. In order to randomize the effects of offset, gain, and timing mismatch a ping-ping architecture [281] uses a third channel. The flash converter digitizes 5 MSBs during a sampling period, while one SAR converts the LSBs of the previous sample. This leaves two SARs of which one is randomly selected to sample the signal and determine the remaining bits, Fig. 9.33. The SAR range is extended with an MSB-4 bit (this is a second MSB of the SAR indicated as "R") to correct errors in the flash conversion. With a sample rate of $f_s = 80$ Ms/s, a power of 31.1 mW, an effective resolution of 11.55 ENOB is obtained in a 65-nm CMOS process.

The design in Fig. 9.34 achieves a 10.3 Gs/s sample rate for 6-bit performance with a four fold time-interleave scheme. In order to avoid interference of the multiplexed T&H circuits, see Fig. 9.8, the input is first buffered in two streams and sampled with opposite phase clocks. After that another buffer stage interfaces

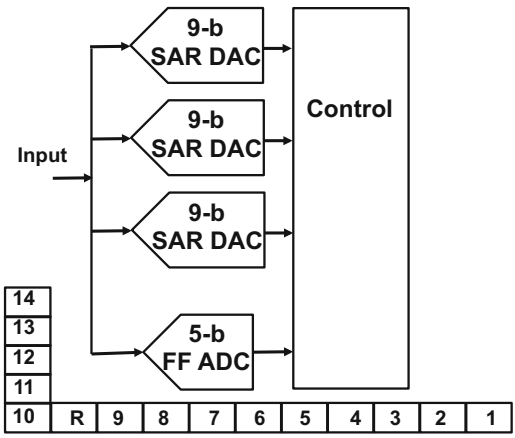


Fig. 9.33 A 5-bit flash ADC takes a first decision after which the one of three 9-bit SARs process the LSBs [281]

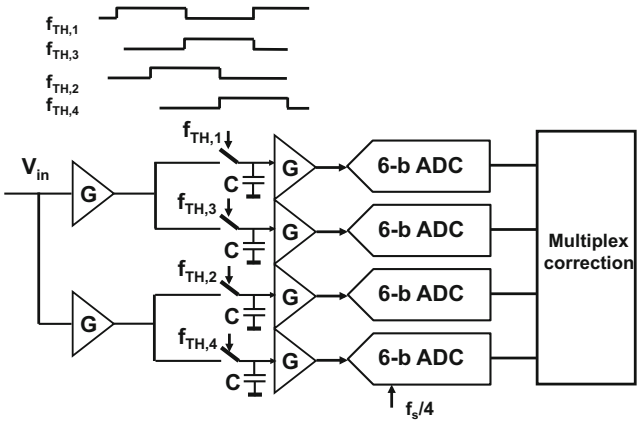


Fig. 9.34 A 10.3 Gs/s four times interleaved flash converter. Simplified from [287]

to the 6-bit flash converters. The flash comparators are calibrated and a Wallace-tree decoding is used. At all stages error correction and calibration is applied. An F.o.M. of 0.7 pJ/conv. step is the obvious efficiency penalty.

9.4.3 High N_{int}

Higher interleaving factors do not necessarily result in faster overall sampling rates. The loading of the driver stage with more sample-and-hold capacitors (to relax the settling) requires a lot of current. Moreover, interference between channels, and the need to use efficient analog-to-digital topologies limit the obtainable performance. Still interleaving by 10, e.g., [254] or 16, e.g., [286] can fit well to the sequencing of the successive approximation converter. Other arguments to go to high interleaving factors are the efficiency of the underlying converters, sufficient decision time for the comparators, and specific demands by the application.

In the architecture of Fig. 9.35 [282] eight flash converters are interleaved to obtain a 12 Gs/s 5-bit resolution conversion system. At the Nyquist frequency any timing skew between the channels would inevitably lead to performance degradation. The design therefore uses a calibration method based on the autocorrelation of the signal in a channel with the sign of the signal sampled at 1/9 fraction of the full sample rate. This fraction causes that the comparator samples one T_s period slower than the eight channels. And that every next comparator sample coincides with the next channel. From the autocorrelation the actual delay information is retrieved. A delay block that tunes the sample pulse per channel to meet the required timing accuracy. The process converges in approximately 100 ms. See also [300] for an signal derivative-based algorithm.

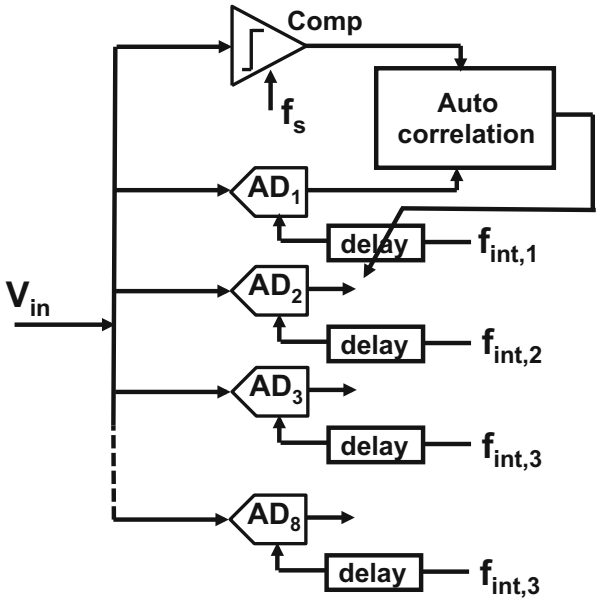


Fig. 9.35 The jitter errors are calibrated via autocorrelation with the sign of the signal [282]

9.4.4 Two-Stage Interleaving

The trade-off between a low and a high interleaving factor is resolved by using a two-stage interleaving approach. The issues with sampling jitter are solved by using one or a few input T&H stages. After a second round of interleaving the total interleave factor is increased to 8,. . . ,64. A intermediate step towards two-stage interleaving is seen in Fig. 9.32, where it is the residue of the MDAC stage (not the full-signal) that is interleaved over four successive approximation converters.

Figure 9.36 shows an architecture with an input T&H circuit running at the maximum sample rate of 2.5 Gs/s with 70 fs_{rms} jitter. The interleave factor of eight has been chosen in order to give the comparators in a 130 nm BiCMOS process sufficient time to reach a BER of 10^{-17}. The structure is chopped and dither is applied. The zero-mean chopping sequence allows to calibrate any DC-offset. And the injected dither is measured so the gain errors can be removed. The distortion level for a 1.052 GHz signal is better than −80 dB, and the total converter consumes some 23.9 W [301].

An earlier publication [288] shows a similar structure: one input T&H circuit followed by a four-time-interleaved pipeline section. After gain and offset calibration the spurs and distortion are at a −70 dB level for 350 mW in 90 nm CMOS.

Figure 9.37 shows the block diagram of a 64-channel two-stage interleaved 3.6 Gs/s successive approximation converter [283, 290]. The signal is sampled and distributed by four first-stage track-and-hold circuits. Here the timing and

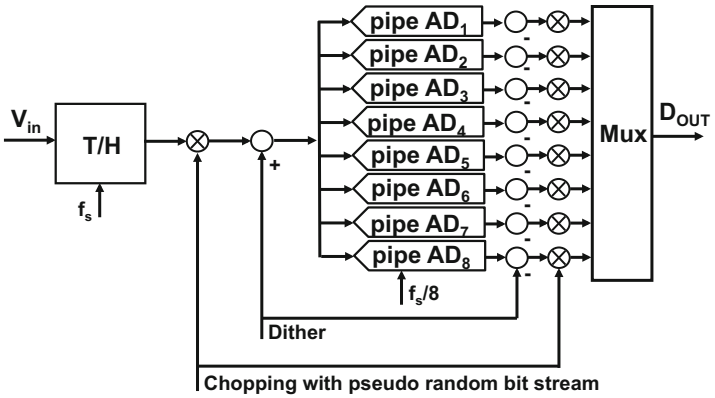


Fig. 9.36 A two-stage interleaving architecture with a single input T&H stage and eight 14-bit pipeline converters. Simplified schematic diagram from [301]

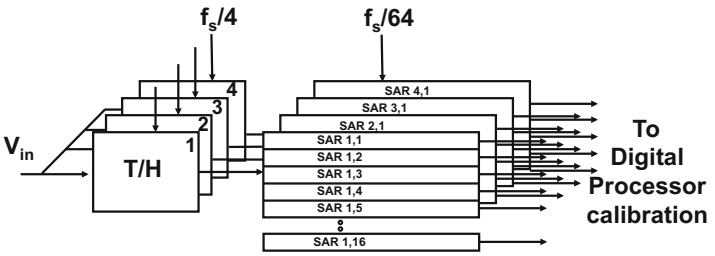


Fig. 9.37 Example of a multiplexed successive approximation analog-to-digital converter, for digitizing a full-spectrum video signal DOCSIS3.0 [283, 290]

bandwidth mismatch is crucial. These circuits each drive the track-and-hold circuits in the four section each containing 16 successive approximation converters. As a settled output signal of the four main track-and-hold circuits is available, the timing problem for the second stage T&H is considerably reduced. Chopping and randomized allocation over the SAR channels reduce the effect of time-interleave errors. Figure 9.38 gives an impression of the lay-out: the main track-and-hold circuits are placed in the center in order to keep the wires to the four sections as balanced as possible. An overall performance of 50 dB for 1 GHz input signals in a 65-nm technology is achieved with a bandwidth $BW = 1\,\text{GHz}$, $f_s = 3.6\,\text{Gs/s}$ for $P = 795\,\text{mW}$ [290]. The overall reported jitter is 0.11 ps_{rms} and timing skew 0.57 ps_{rms}.

In Fig. 9.39 some of the calibration mechanisms are depicted. The central background calibration averages out the remaining offsets. For gain calibration the power average of the channels is compared.

The effect of the calibration is visible in Fig. 9.40.

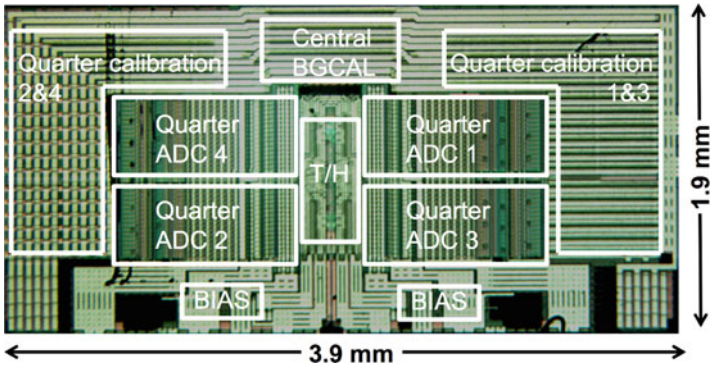


Fig. 9.38 Chip photograph of a multiplexed successive approximation analog-to-digital converter [283, 290]. Courtesy photo: NXP design team

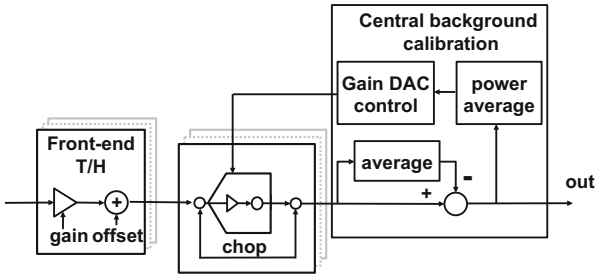


Fig. 9.39 Simplified view of the offset and gain calibration [290]. Courtesy NXP design team

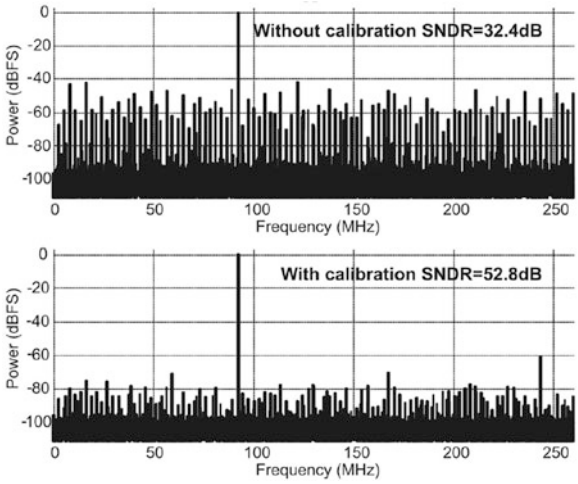


Fig. 9.40 Comparison of spurs with and without calibration [290]. Courtesy NXP design team

In [284] a similar architecture as in Fig. 9.37 was pushed in 32-nm SOI CMOS to 90 Gs/s speed performance. With a power of 667 mW an input signal of 19.9 GHz was converted at 5.2 ENOB.

9.5 Frequency Multiplexing

The previous sections have extensively elaborated on the ideas for demultiplexing a signal stream in the time domain. In fact the idea of demultiplexing in the frequency domain is much older as it is the basis for any radio transmission. A 100 GHz real-time bandwidth oscilloscope (LabMaster 10–100zi, LeCroy) applies both ideas in one data acquisition front-end, Fig. 9.41. Special microwave filters split the initial signal in three bandwidths. The signal components inside two higher bandwidths are brought to a lower frequency band by means of a mixer section. The 16-times interleaved track-and-hold circuit performs a time-demultiplex and prepares the conversion in 8-bit samples. According to specification the internal sample clock jitters at $\sigma_{Ts} = 50\,\text{fs}$, resulting in an SNR_{jitter} of 46 dB. Finally a large digital processing unit performs the necessary corrections and re-assembles the signal for rendering it onto a display.

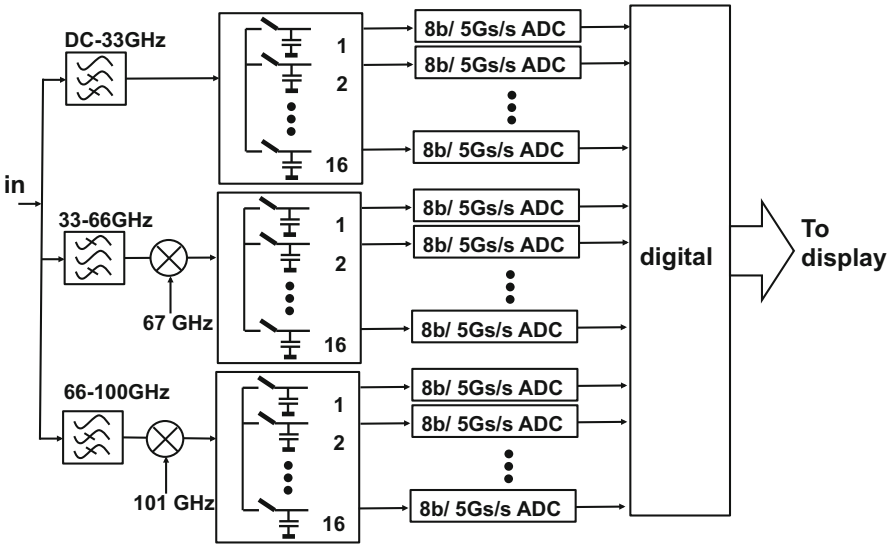


Fig. 9.41 The architecture for a 100 GHz bandwidth, real-time, 240 Gs/s oscilloscope front-end (LeCroy LabMaster), after a description on: www.thesignalpath.com

The system is partitioned into special microwave filters, a proprietary T&H chip and a multitude of analog-to-digital converter chips.

9.6 A Comparison

Time-interleaving is exploring new fields in the analog-to-digital conversion area. Until a few years ago high-speed converters were designed as flash or pipeline converters. Now time-interleaving has gain a clear position in that map, Fig. 9.42. Pipeline converters cluster between 50 and 80 dB (8–13 ENOB) of SINAD at sample-rates up to 300 Ms/s, while flash converters make up the 4–5 ENOB field. Time-interleaving allows to raise the performance of the flash converters to almost 6 ENOB at similar speeds. And the lower SINAD (50–60 dB) range of the pipeline converters is overtaken by time-interleaved (pipeline) converters.

Still the inevitable barrier of channel inequalities seems to bar the 10+ ENOB range above $f_s = 100$ Ms/s.

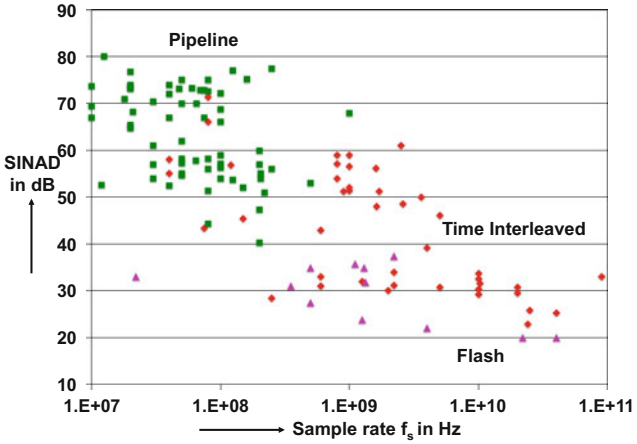


Fig. 9.42 Comparison of time-interleaved converters to pipeline and flash topologies (Data from: B. Murmann, ADC Performance Survey 1997–2015, Online: http://web.stanford.edu/?murmann/adcsurvey.html)

Exercises

9.1. A 64-time-interleaved converter shows an offset on the channels of $\sigma_{off} = 1\,\text{mV}$, a gain variation of 0.1 %, a sampling skew of $1.5\,\text{ps}_{rms}$, and a bandwidth variation of 0.5 % where the bandwidth is equal to f_s and the maximum signal (top-top) $1\,\text{V}_{pp}$. For what signals can the best performance be reached and what is that performance (SNR)?

9.2. As in the previous exercise: at what input frequency and amplitude is the SNR loss due to the static mismatch (offset and gain) equal to the dynamic (sample skew and bandwidth variation)?

9.3. Compare the architecture of the designs in [290] and [284]. What is the main difference?

9.4. Explain why a sigma-delta converter is not suited for time-interleaving.

Chapter 10
Sigma-Delta Modulation

The various Nyquist conversion principles discussed in the previous chapters allow to realize analog-to-digital converters for broadband signals. Many applications do not require wide-band conversion as the bandwidth is limited, such as in various communication systems. Moreover, in advanced CMOS processes GigaHertz-range sampling and processing frequencies are at the disposal of the designer. These high speeds do not fit to the needs of many applications. Oversampling, noise shaping, and sigma-delta modulation form a different class of principles for analog-to-digital conversion of limited bandwidth signals in technologies with a high sample rate. De Jager [302], Cutler [303], and Widrow [304] were the first researchers to propose the idea that in a limited bandwidth a reduced accuracy per sample can be compensated by using a high sample rate. Inose [305] proposed an architecture featuring the filter operation in the loop, thereby creating today's sigma-delta modulator topology.

In the next sections the various steps towards efficient narrow band analog-to-digital conversion will be discussed: oversampling, noise shaping, and sigma-delta conversion.

10.1 Oversampling

10.1.1 Oversampling in Analog-to-Digital Conversion

In the process of sampling and quantization an error signal is generated that is fundamentally a distortion component. This quantization error and the multitude of distortion components folded back by various multiples of the sampling frequency are approximated as white noise as long as the resolution N is sufficiently large and no correlation appears between the input signal and the sample rate (one frequency is not an integer ratio of the other). This quantization error has a fixed amount of

M. Pelgrom, *Analog-to-Digital Conversion*, DOI 10.1007/978-3-319-44971-5_10

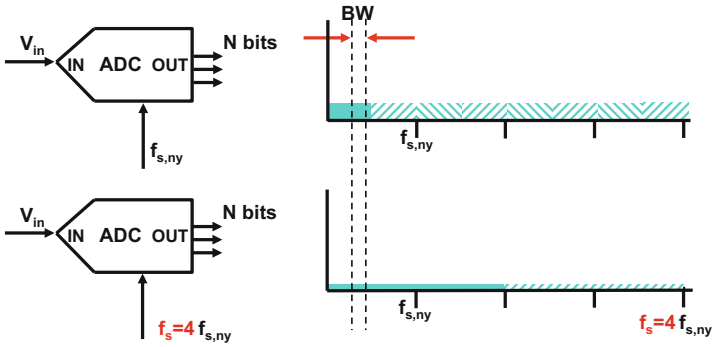


Fig. 10.1 Oversampling in analog-to-digital conversion. In the derivations and the next figures the lower boundary of the bandwidth is set to 0 Hz. This is not a necessary condition but eases the math

power $A_{LSB}^2/12$ for a given resolution N and an LSB of A_{LSB}. The error spectrum is modeled as a noise-like component that is uniformly ("white") distributed over the band from 0 to $f_s/2$ and mirrored into the alias bands.

If the sample rate of an analog-to-digital converter is increased from the frequency needed to fulfill the Nyquist criterion $f_{s,ny} > 2BW$ to a new frequency f_s the noise power density (power per Hertz) is reduced with the ratio of the sample rates. Figure 10.1 shows the increase of the sampling rate by a factor of four. As the noise power density is reduced by a factor of four, the amplitude noise density is reduced by a factor of two. In a fixed bandwidth the signal-to-noise ratio will increase by this factor of four in power density and the effective resolution (ratio between signal and unwanted components) increases by one effective bit. The ratio

$$\text{OSR} = \frac{f_s}{f_{s,ny}} = \frac{f_s}{2f_b} \tag{10.1}$$

is called the oversampling ratio (OSR). The bandwidth BW of the signal is for ease of calculation assumed to range from 0 to f_b. The same result is obtained for bands at other frequency locations.

In a Nyquist converter the quantization power $Q^2 = V_{LSB}^2/12$ is supposed to be uniformly distributed over the frequency band till half of the sample rate (and repeats in the next $f_s/2$-wide sections). The corresponding noise power density is equal to the total noise divided by half of the sampling rate $f_s/2$:

$$N_Q(f) = \frac{Q^2}{f_s/2} = \frac{V_{LSB}^2/12}{f_s/2} = \frac{V_{LSB}^2}{6f_s} \tag{10.2}$$

In a band from 0 to f_b the total noise power Q_b^2 is found by integrating the noise density over the band from 0 to f_b resulting in a noise power:

$$Q_b^2 = \frac{V_{LSB}^2 f_b}{6 f_s} = \frac{V_{LSB}^2}{12} \frac{1}{\mathrm{OSR}} \tag{10.3}$$

The above relation will be used to compare the upcoming results for noise shaping. The noise power in a fixed bandwidth reduces proportionally to the oversampling rate. The gain in signal-to-noise ratio and in effective number of bits in a fixed bandwidth is

$$\Delta\mathrm{SNR} = {}^{10}\log(\frac{f_s}{f_{s,ny}}) = {}^{10}\log(\mathrm{OSR}) \tag{10.4}$$

$$\Delta\mathrm{ENOB} = \frac{1}{2}{}^{2}\log(\frac{f_s}{f_{s,ny}}) = \frac{1}{2}{}^{2}\log(\mathrm{OSR}) \tag{10.5}$$

Oversampling is not effective for signals where the assumption that the quantization process can be approximated by white noise is not valid. The signal-to-noise ratio of the conversion of DC-signals cannot be improved by oversampling alone. A helper signal must be present to create the necessary conditions, e.g., in the form of dither, see Sect. 4.3.4.

The main advantage of oversampling in analog-to-digital conversion is the extra frequency range that is obtained between the band of interest and the alias band, see Fig. 2.9. The specifications of the anti-alias filter can thereby be relaxed, although the higher sample rate will lead to higher power consumption on the digital side.

10.1.2 Oversampling in Digital-to-Analog Conversion

In a digital-to-analog conversion it is even more beneficial to increase the sample rate and avoid an operation mode close to the Nyquist limit. Figure 10.2 shows an example of a digital-to-analog conversion of a bandwidth close to its Nyquist limit. The digital-to-analog converter is succeeded by an off-chip filter for suppressing the alias band. At signal frequencies close to half of the sample rate, large transient steps will occur at the input of the on-chip buffer and the output pin. These steps cause

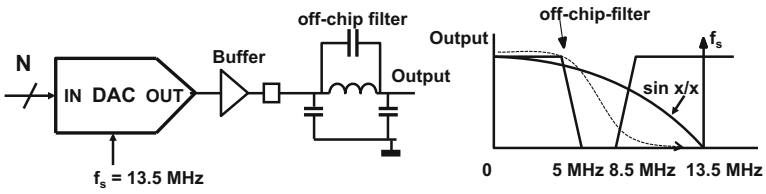


Fig. 10.2 A digital-to-analog converter in a video application converting a signal bandwidth close to the Nyquist rate. When the output aliases are filtered by an external filter, the requirements on the filter and the buffer become hard to meet

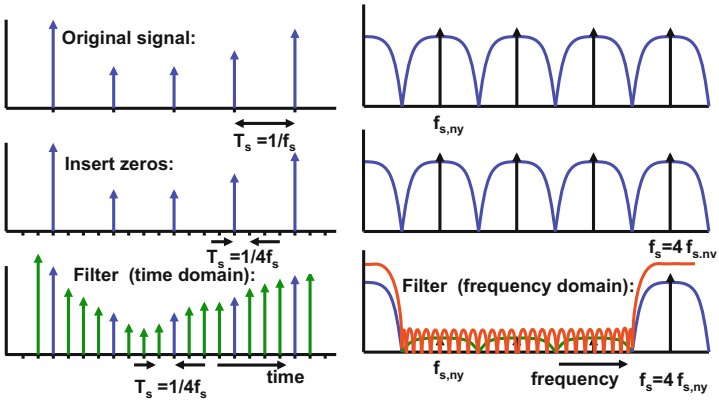


Fig. 10.3 Four-fold digital oversampling in time and frequency domain

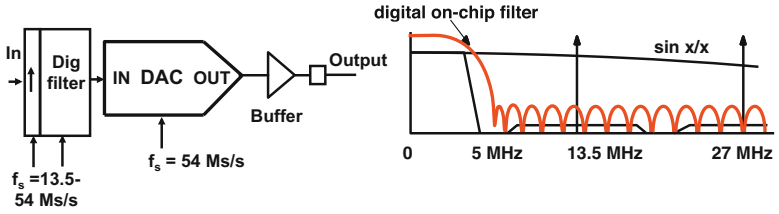


Fig. 10.4 A digital-to-analog converter driven by digital pre-filter that performs a two-time oversampling

slewing and distortion in the buffer. The buffer has to be designed with additional bandwidth and the input stages will need large bias currents. This set-up has a relatively poor performance close to $f_s/2$ due to sin(x)/x signal loss. The passive filter requires some 3–7 poles and is expensive to produce, especially if sin(x)/x compensation is also needed.

The solution to this problem is to construct in the digital domain a signal at a higher sampling rate: oversampling. In order to create an oversampled version of the digital signal, a number of processing steps must be taken, see Fig. 10.3. Between the original samples at f_s new sample points are inserted with a value "0", in the example at $t = T_s/4, 2T_s/4, 3T_s/4, \ldots$. Although the frequency spectrum is still the same, this step causes the hardware to run at $4f_s$, which is formally the new sample rate. Now a digital filter (see Sect. 2.7.1) removes the alias bands. In the time domain the filter operation will change the value of the inserted sample points to new values. Some authors refer to this method as "up-sampling" in order to emphasize the contrast with subsampling of a sigma-delta output signal described in Sect. 2.3.3.

Figure 10.4 shows an integrated circuit solution: the sample rate is locally doubled and the alias bands in the frequency spectrum are suppressed by digital

Table 10.1 Comparison of the two digital-to-analog conversion solutions for a video application in Figs. 10.2 and 10.4

Nyquist rate solution	Oversampling solution
External multi-pole filter needed	Internal digital CMOS filter
High slew current in driver	Medium current in driver power for digital filter
sin(x)/x loss of 2 dB	sin(x)/x loss = 0.3 dB
Standard sample rate	4× sample rate needed

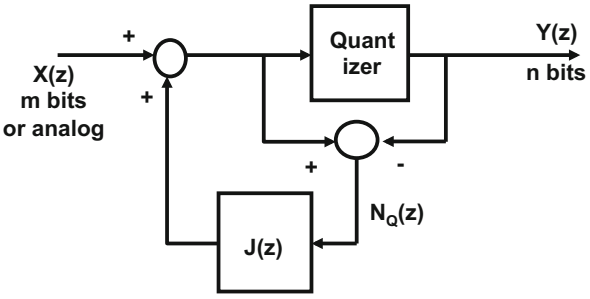


Fig. 10.5 Signal-to-noise improvement with noise shaping can be applied on analog or rounding of m-bit digital signals, where $m > n$

filtering. Now the large transients in the output are more than halved in amplitude, and relatively simple non-critical post-filtering (first order) is sufficient to restore the analog signal. The inherent sin(x)/x at 5 MHz is reduced from −2 to −0.5 dB, so no compensation is required. From a power perspective a trade-off must be made between the additional power and area for the filter and the quality loss and additional power in the buffer (see Table 10.1). In high-speed and efficient CMOS processes this trade-off favors the oversampling solution. In advanced CMOS technology the area and power of the digital filter shrinks and the switching speed of the short channel transistors allows high oversampling frequencies.

10.2 Noise Shaping

Oversampling is a useful concept for relaxing the alias filter requirements and buffer specifications in the total conversion chain. A second reason for applying oversampling is that it creates an additional frequency span. This additional frequency range is used in noise shaping for shifting the quantization power out of the wanted signal band into a part of the frequency span where it no longer affects the signal.

Figure 10.5 shows the basic structure of a noise shaper circuit. The quantization error, or quantization "noise," is formed by subtracting the input and output signals

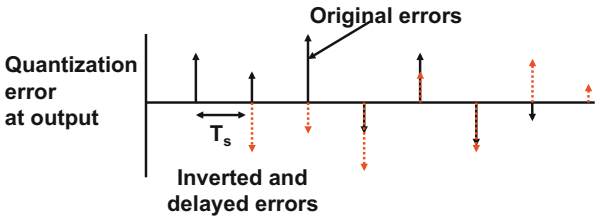


Fig. 10.6 The output of the noise shaper contains both the original error sequence and its inverted and delayed version. Both are not visible at the output as in this plot, because they are part of the quantized signal. This plot helps in realizing that the DC-content of the error combination is low

from the quantizer.[1] This quantization error is passed through a filter $J(z)$ and fed back to the input. The inverted quantization errors are delayed and added to the signal, see Fig. 10.6. Low-frequency components in the error signal are effectively suppressed. At high frequencies the filter delay will cause a phase shift leading to an enhanced error power level. This structure is not a classical feedback loop. There is no signal that is fed back to its own origin. Therefore there is no feedback stability problem, which opens a wide range of choices for $J(z)$.

In the z-domain the transfer of input and noise into the output is

$$Y(z) = X(z) + [1 - J(z)]N_Q(z) \tag{10.6}$$

If the filter function $J(z)$ is chosen as a unit delay:

$$Y(z) = X(z) + [1 - z^{-1}]N_Q(z) \tag{10.7}$$

This transfer function can be visualized in the frequency domain via the transformation $z \leftrightarrow e^{j\omega T_s}$:

$$Y(\omega) = X(\omega) + [1 - e^{-j\omega T_s}]N_Q(\omega) \tag{10.8}$$

In order to determine the noise transfer to the output (NTF), the absolute value of the function in brackets is evaluated:

$$\begin{aligned} |\mathrm{NTF}(\omega)|^2 &= \left|\frac{Y(\omega)}{N_Q(\omega)}\right|^2 = \left|1 - e^{-j\omega T_s}\right|^2 \\ &= 2 - 2\cos(\omega T_s) = 4\sin^2(\omega T_s) = 4\sin^2(\pi f/f_s) \end{aligned} \tag{10.9}$$

[1] As mentioned in Chap. 4 quantization errors specifically of low-resolution quantizers are not noise, but folded distortion products. With the introduction of the "white noise" model for quantization errors, the incorrect terms "noise" and "noise shaper" have become an accepted description.

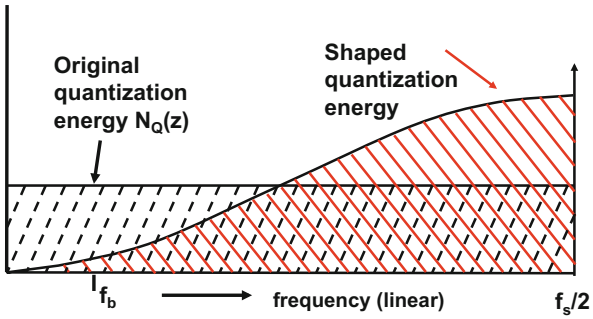


Fig. 10.7 The noise shaper pushes the quantization power to higher frequencies

In the noise shaper the quantization power density is shaped with the function $(2 - 2\cos(\omega T_s))$. Figure 10.7 shows the resulting noise power density for the filter function $J(z) = z^{-1}$. Integration of the noise power density function over the band from 0 to $f_s/2$ gives

$$\frac{V_{LSB}^2}{6f_s} \int_{f=0}^{f=f_s/2} (2 - 2\cos(2\pi f/f_s))\, df = 2 \times \frac{V_{LSB}^2}{12} \tag{10.10}$$

which equals twice the quantization noise power.

This can be understood by considering that the amplitudes of succeeding quantizations are uncorrelated. The extracted quantization errors are delayed and combined with the present quantization error to yield the output value, see Fig. 10.6. Now two (in amplitude) uncorrelated quantization sequences appear at the output, the total quantization error power doubles, although it is shaped in the frequency domain.

For small values of the cosine argument the approximation $\cos(x) \approx 1 - x^2/2$ results in a noise power density:

$$\frac{V_{LSB}^2}{6f_s}(2 - 2\cos(\omega T_s)) \approx \frac{V_{LSB}^2}{6f_s}\left(\frac{2\pi f}{f_s}\right)^2 \tag{10.11}$$

The noise power density near DC is strongly reduced. The quantization noise amplitude increases with $2\pi f/f_s$ which corresponds to a first order frequency behavior. Close to $f_s/2$ the power density is four times the Nyquist power density, so the noise shaper does not eliminate the noise but shifts it to a different frequency region.

The total noise power in a band from 0 to f_b in the noise shaper is found by integrating the noise power density over that frequency range:

$$\int_{f=0}^{f_b} \frac{V_{LSB}^2}{6f_s}|1-e^{-2j\pi fT_s}|^2 df = \frac{V_{LSB}^2}{3}\left[\frac{f_b}{f_s} - \frac{\sin(2\pi f_b/f_s)}{2\pi}\right] \approx$$

$$\frac{V_{LSB}^2}{12}\frac{\pi^2}{3}\left(\frac{2f_b}{f_s}\right)^3 = \frac{V_{LSB}^2}{12}\frac{\pi^2}{3\text{OSR}^3} \tag{10.12}$$

The total in-band noise of a noise shaper with respect to only oversampling conversion is found by dividing this result by the total in-band noise of an oversampling conversion, Eq. 10.3:

$$\text{Noise power reduction} = \left[2 - 2\frac{\sin(2\pi f_b/f_s)}{2\pi f_b/f_s}\right] \tag{10.13}$$

Approximating the sine with $\sin(x) = x - x^3/6$:

$$\text{Noise power reduction} = \frac{4}{3}(\pi f_b/f_s)^2$$

$$\text{Noise amplitude reduction} = \frac{2}{\sqrt{3}}(\pi f_b/f_s) \tag{10.14}$$

The total in-band noise power is related to the oversampling ratio OSR via a cube power in Eq. 10.12. One OSR term comes from the oversampling mechanism, the remaining OSR^2 term comes from the noise shaping loop. Doubling the oversampling ratio reduces the noise power by a factor 8, or 9 dB or 1.5 effective number of bits. Thereby noise shaping is a powerful mechanism to improve the effective resolution of a converter.

Example 10.1. In a noise shaper the feedback path function equals $J(z) = z^{-2}$. Sketch the noise behavior at the output of this noise shaper.

Solution. Figure 10.8 shows the noise spectrum in the output. In case $J(z) = z^{-2}$, the shape of the noise in the output is $1 - J(z) = 1 - z^{-2}$, which can be transformed to the frequency domain via $z \leftrightarrow e^{j2\pi f}$.

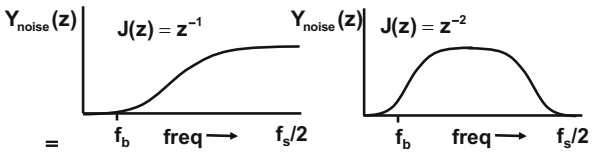


Fig. 10.8 *Left*: the noise spectrum in the output with a single sample pulse delay. *Right*: the spectrum with two sample pulses delay

10.2.1 Higher-Order Noise Shaping

Next to first order noise shaping, also higher-order shaping is possible by extending the filter $J(z)$. If this function is chosen as an n-th order polynomial:

$$1 - J(z) = (1 - z^{-1})^n \tag{10.15}$$

$$Y(z) = X(z) + (1 - z^{-1})^n N_Q(z) \tag{10.16}$$

The frequency transfer function becomes

$$Y(\omega) = X(\omega) + (1 - e^{-j\omega T_s})^n N_Q(\omega) \tag{10.17}$$

In order to determine the noise transfer to the output (NTF), the absolute value of the function in brackets is evaluated:

$$\begin{aligned} |\mathrm{NTF}(\omega)|^2 &= \left|\frac{Y(\omega)}{N_Q(\omega)}\right|^{2n} = \left|1 - e^{-j\omega T_s}\right|^{2n} \\ &= (2 - 2\cos(\omega T_s))^{2n} = (2\sin(\pi f/f_s))^{2n} \end{aligned} \tag{10.18}$$

Figure 10.9 shows the evaluation of this function for $n = 1, 2$. Now the suppression of low-frequency quantization components is considerably improved.

This is certainly not the only possible higher-order filter for a noise shaper. However this choice results in a manageable mathematical description, showing the impact of the additional order of the filter. The total noise power in a band from 0 to f_b in an n-th order noise shaper is found in a similar way as the first order[2]:

$$\int_{f=0}^{f_b} \frac{V_{LSB}^2}{6f_s} |1 - e^{-2j\pi f T_s}|^{2n} df = \int_{f=0}^{f_b} \frac{V_{LSB}^2}{6f_s} (2\sin(\pi f/f_s))^{2n} df$$

$$= \frac{V_{LSB}^2}{6\pi} \frac{(2n)!}{(n!)^2} \left[\pi \frac{f_b}{f_s} - \cos\left(\pi \frac{f_b}{f_s}\right) \sum_{k=0}^{k=n-1} \frac{2^{2k}(k!)^2 (2n)! \sin^{2k+1}(\pi f_b/f_s)}{(2k+1)!} \right] \tag{10.19}$$

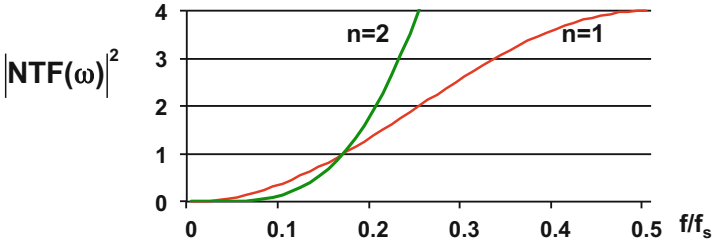


Fig. 10.9 The power noise transfer function is evaluated for $n = 1, 2$. Close to DC the suppression of quantization power is considerably enhanced

[2]Formula [8, formula 300] is here useful as pointed out by V. Zieren.

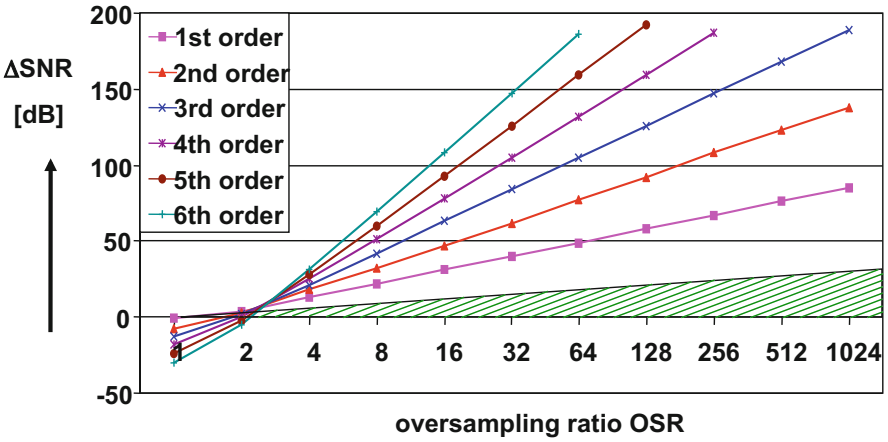


Fig. 10.10 The gain in signal-to-noise ratio for an n-th order noise shaper as calculated in Eq. 10.20. The *shaded triangle* is the portion that originates from oversampling alone. This plot is equally valid for sigma-delta conversion. The same *curves* apply to n-th order sigma-delta conversion

The signal-to-noise ratio gain is found by dividing this result by the total noise in the Nyquist band $V_{LSB}^2/12$:

$$\text{SNR gain} = \frac{2}{\pi}\frac{(2n)!}{(n!)^2}\left[\pi\frac{f_b}{f_s} - \cos\left(\pi\frac{f_b}{f_s}\right)\sum_{k=0}^{k=n-1}\frac{2^{2k}(k!)^2(2n)!\sin^{2k+1}(\pi f_b/f_s)}{(2k+1)!}\right] \tag{10.20}$$

Figure 10.10 shows the signal-to-noise gain for first till sixth order noise shaping evaluating this formula.

Assuming that f_b/f_s is sufficiently small, a simpler expression can be obtained for the noise power:

$$\int_{f=0}^{f_b}\frac{V_{LSB}^2}{6f_s}|1-e^{-2j\pi f T_s}|^{2n}df \approx \int_{f=0}^{f_b}\frac{V_{LSB}^2}{6f_s}(2\pi f/f_s)^{2n}df$$

$$= \frac{V_{LSB}^2}{12\pi}\frac{(2\pi f/f_s)^{2n+1}}{2n+1} = \left(\frac{V_{LSB}^2}{12}\right)\times\left(\frac{1}{OSR}\right)\times\left(\frac{\pi^{2n}}{(2n+1)\text{OSR}^{2n}}\right) \tag{10.21}$$

In the last formulation the different processes have been made explicit: the quantization power, the signal-to-noise gain by mere oversampling, and the signal-to-noise gain by noise shaping. The noise power reduces by $(2n+1)\times 3\,\text{dB}$ for every doubling of the oversampling rate. The SNR improvement of n-th order noise shaping over Nyquist conversion is given as:

$$\Delta\text{SNR} = 10^{10}\log\left((2n+1)\frac{\text{OSR}^{2n+1}}{\pi^{2n}}\right) = 20^{10}\log\left(\frac{\text{OSR}^{n+0.5}\sqrt{2n+1}}{\pi^{n}}\right) \tag{10.22}$$

This expression has an error of less than 1 dB with respect to Eq. 10.20 and Fig. 10.10 for oversampling ratios larger than 4.

This expression is based on the assumption in Eq. 10.15 that $1 - J(z)$ is a Butterworth filter. Other filters lead to different curves. Yet, this analysis gives a first impression on what improvement can be reached.

Noise shaping has been demonstrated here on abstract input and output signals. Often noise shaping is considered for quantizing analog signals into the digital domain. Due to the change of domain around the noise shaper, the digital output signal must then be converted back to the analog domain before the subtraction from the analog input signal can take place. In case of a gain error the output signal Y is (slightly) attenuated by $(1 - c)$ and the transfer becomes

$$Y(z) = \frac{X(z)}{1 - cz^{-1}} + \frac{1 - J(z)}{1 - cz^{-1}}Q(z) \tag{10.23}$$

which implies a linear signal distortion for the input signal and an less accurate filter function for the quantization error. This application of the noise shaping idea requires some tuning of the analog signals, in order to obtain the correct transfer function.

Example 10.2. A digital signal processor delivers a 10-bit result. This result must be processed to fit optimally to an 8-bit bus. Give a solution using noise shaping.

Solution. The problem to truncate a 10-bit digital data word to 8 bits can be solved by means of a simple form of a first order noise shaper as shown in Fig. 10.11. The truncation is a form of quantization. The two residue bits form the quantization error. These two bits are fed back via a delay and are added into the original signal. The DC-content of the signal is preserved, as all bit information is added together. However the consequence is that the carry-bit due to the successive addition of the

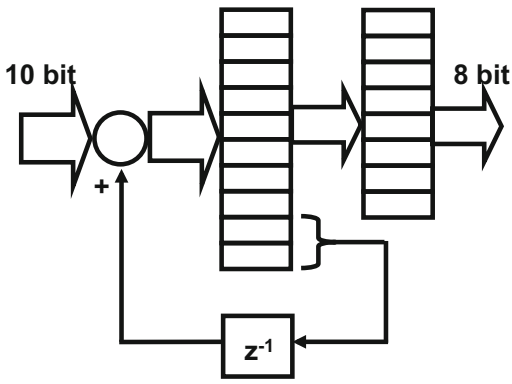


Fig. 10.11 Noise shaping in digital rounding

9th and 10th bit occurs at irregular intervals in the new output word. Giving some high-frequency noise. The overall quantization/truncation error at $t = nT_s$ is now:

$$D_{out}(nT_s) = D_{in}(nT_s) - Q(nT_s) + Q((n-1)T_s)$$

The output equals the input minus the truncation error plus the previous error. The quantization error is therefore subject to signal processing in the z-domain:

$$Q_{out}(z) = (1 - z^{-1})Q(z)$$

With the substitution $z \Leftrightarrow e^{j\omega T_s}$ it is clear that the quantization noise is first order shaped with a zero at DC. Compare the VCO-based quantizer in Fig. 8.127 [273].

10.3 Sigma-Delta Modulation

An important solution for converting a low-bandwidth analog signal into a low bit-width pulse stream is the sigma-delta[3] converter[4] as shown in Fig. 10.20, e.g., [305–308].[5] Recent overviews of this field are presented in [309, 310]. In this Sect. 10.3 a global overview of sigma-delta modulation is given. The detailed topologies and implementation aspects are treated in the next sections.

10.3.1 A Quantizer in a Feedback Loop

Sigma-delta conversion as a form of signal quantization can be used in analog-to-digital, digital-to-analog, and digital-to-digital conversions of signals. In contrast to the noise shaper, both the signal component and the quantization error are circulated in the feedback path of the sigma-delta converter. The stability of the loop is the critical design challenge. The quantizer can vary from a simple comparator to a flash converter, successive approximation of "digital" quantizer such as oscillators. In this scheme the quantizer operates on the filtered residue signal and reduces this residue signal to a discrete amplitude signal.

[3]Is the term "sigma-delta" or "delta-sigma" more correct? Inose [305] uses in 1962 delta-sigma. The opposing argument is that the basic form was originally a delta-modulator which was extended with an summing function: a sigma-delta modulator.

[4]More language issues: "modulator" or "converter." In this book the circuitry around the quantizer is referred to as the modulator. The same circuit is called a converter if its system function is the dominant feature.

[5]Many authors have published books on sigma-delta modulation, still the ultimate text has not been written. Books on this subject tend to be too mathematical, too magical, or both. First spend an hour with your upcoming purchase, before you decide! Also recommended for this book.

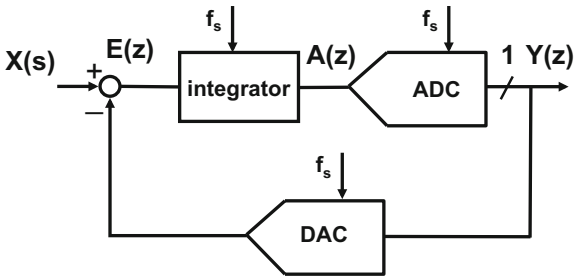


Fig. 10.12 Basic scheme of a sigma-delta modulator

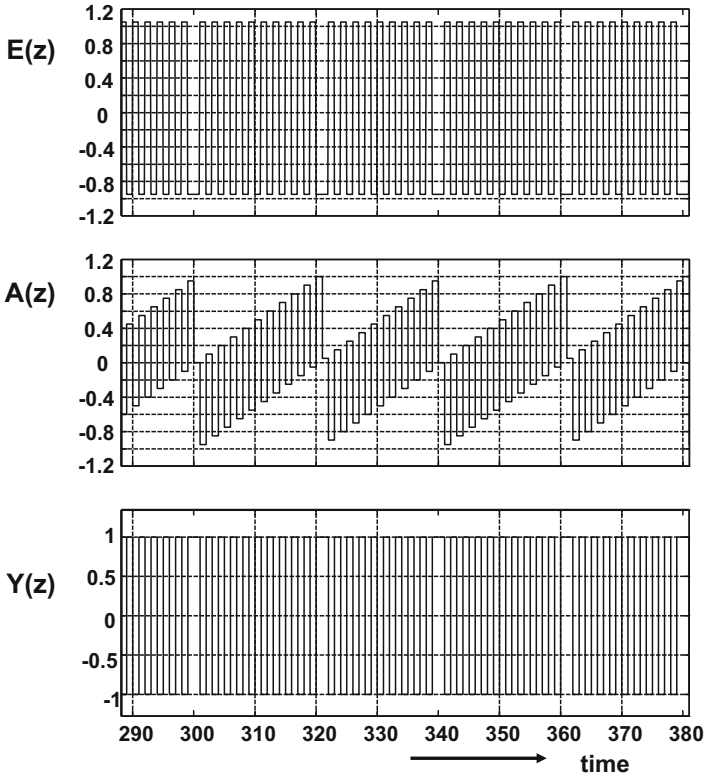


Fig. 10.13 The signals in a first order time-discrete sigma-delta modulator with a small DC input signal. *Top*: the signal after the summation node. *Middle*: the integrator output. *Bottom*: the comparator output

Figure 10.12 shows a first order time-discrete sigma-delta modulator. The loop filter is implemented as a first order time-discrete integrator. In the simulation shown in Fig. 10.13 the dominant signal in the loop is a block wave with frequency components of $f_s/2$. The sigma-delta modulator produces a stream of alternating bits even if the input signal is a DC-level. This block wave at the output of the modulator

is inverted when it passes the input summation node and is then integrated on the next clock edge. The resulting signal provides the input for the comparator. The combination of the clocking of the integrator and the inversion in the loop provide sufficient conditions to implement a stable oscillation at half of the switching rate.

In Fig. 10.13, a DC signal is superimposed on this oscillation via the input terminal. The DC signal causes the output of the summation node $E(z)$ to be shifted a bit positive. As a result the integrator increases slightly its output value on each clock cycle. After repeated passes of the oscillation signal, the filter output reaches a level where the negative-going block wave is not sufficient to reach the threshold of the comparator. The comparator generates a double-length positive pulse, and the oscillation wave skips one transition. In the long-term output these stretched pulses create (after averaging) the correct representation of the DC-input signal. The input signal modulates the oscillation and the ratio of the positive and negative pulses corresponds to the ratio of the input signal to the amplitude range of the digital-to-analog converter: sigma-delta modulation is a form of pulse-density coding. The ratio between positive and negative pulses follows the DC level of the input. A high DC-level in the integrator must be balanced by a large number of positive output pulses that cause negative pulses in the integrator. If the DC-level reaches the maximum amplitude of the feedback digital-to-analog converter, the modulator will just provide positive pulses and is in overrange or "unstable." The same reasoning holds also for large negative inputs.

Note here that a sigma-delta loop essentially oscillates. In control theory the loop is unstable. The design of the sigma-delta loop requires to build an oscillator, for which the well-known Barkhausen criteria apply:

- The round-going loop amplification is exactly 1. Not more, not less.
- The round going phase shift is exactly 360°. Not more, not less.
- For a stable oscillation both conditions must be met.

In a first order sigma-delta converter, the amplification is normalized by the quantizer and digital-to-analog converter: its output amplitude is fixed, and variations in filter output are canceled. The phase of the quantizer output is shifted 180° by the inversion and 90° by the integrator. The remaining phase shift is lost in waiting time for the next quantizer decision. The oscillation is therefore not free running but synchronized to the quantizer's clocking.

The observation that a sigma-delta modulator has some similarity with a voltage-controlled oscillator is helpful in understanding some aspects of the modulators behavior. Another observation is that in a sigma-delta modulator the quantizing error is used as a form of dither to create more resolution for relatively low-frequency signals. Both ways of describing a sigma-delta modulator have their virtues and limitations.

Figure 10.13 reveals also another phenomena in sigma-delta modulation. The lower trace shows an output pattern with a dominant $f_s/2$ component. The regular interruption of this pattern caused by the DC-offset creates low-frequency signal components. These frequency components are called: "idle tones." At low-signal levels these tones can pop up in the desired frequency band [311]. Idle tones

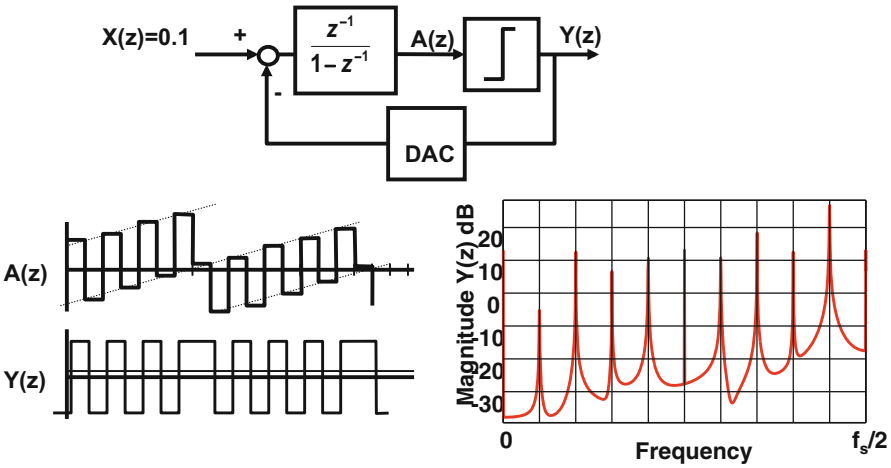


Fig. 10.14 A small DC-offset in a first order sigma-delta modulator creates a pattern with frequency components in the signal band. These idle tones occur in low-order sigma-delta modulators

are a major draw back of sigma-delta modulators. These correlated components, see Fig. 10.14, can present serious problems in applications.[6] With higher-order modulators or the addition of helper signals (dither) these tones are reduced to acceptable levels.

10.3.2 Linearization

As the quantizer is a non-linear element, the modulator cannot be fully captured in a linear system. In a linear model description the quantizer has to be approximated by linear components. Figure 10.15 shows that the quantizer is replaced by a gain factor c and the quantization errors N_Q have been made explicit by adding them with a summation node. With the quantizer's effective gain modeled as c and the digital-to-analog converter gain as d, the input–output relation becomes[7]:

$$Y(z) = \frac{cG(z)}{1 + cdG(z)} X(z) + \frac{1}{1 + cdG(z)} N_Q(z) \tag{10.24}$$

[6] In the audio range these components are audible and are called "whistles." Trained listeners can hear idle tones down to 100 dB below full-signal.

[7] At this introductory point of the description, the variable z is used implying time-discrete behavior. In the following sections a more clear distinction between time-discrete and time-continuous implementations is made.

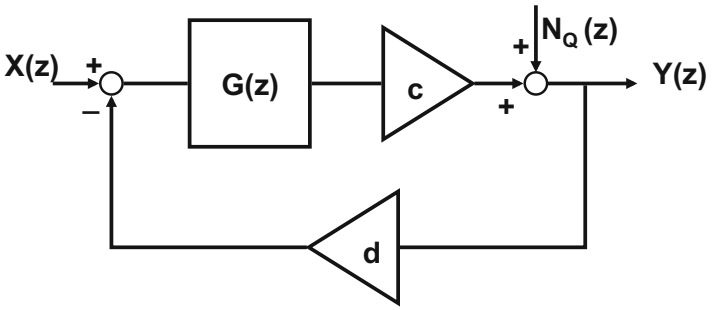


Fig. 10.15 The non-linear quantizer has been replaced by a gain factor and a noise injection point

$G(z)$ is designed to have a high gain within the desired bandwidth of $X(z)$. For those frequencies the equation reduces to:

$$Y(z) = X(z) + \frac{1}{cdG(z)} N_Q(z) \tag{10.25}$$

The feedback loop shapes the quantization error power from a flat spectrum into a shape determined by $1/G(z)$ which is the inverse function of the loop filter. Where the filter creates high-gain, the loop can suppress the quantization errors. In filter theory: the poles of the filter translate in zero's for the noise transfer to the output. The art of sigma-delta design therefore requires designing the right filter to suppress the quantization noise at the desired frequencies.

If this equation is compared to the derivation for the noise shaper, Eq. 10.6, then a mathematical equivalence will occur if:

$$cdG(z) \leftrightarrow \frac{1}{1 - J(z)} = \left(\frac{1}{1 - z^{-1}} \right)^n$$

The last term can be implemented as a cascade of time-discrete integrators. The transfer function of a sigma-delta modulator with a loop filter dominated by an n-th order filter is mathematically comparable to an n-th order noise shaper. The same signal-to-noise improvement as in Fig. 10.10 is obtained.

The gain factor c of one-bit quantizers is created by the comparator. The output of the comparator is given $(-1, +1)$, however, the input signal can be very small. Consequently a rather large gain can be expected. In a design the gain is mostly established from simulating the input power versus the output power, see also Sect. 10.5. For multi-bit quantizers, Sect. 10.7, the ratio between the quantizer LSB size V_{LSB} and the corresponding digital-to-analog converters output step is a good measure for c.

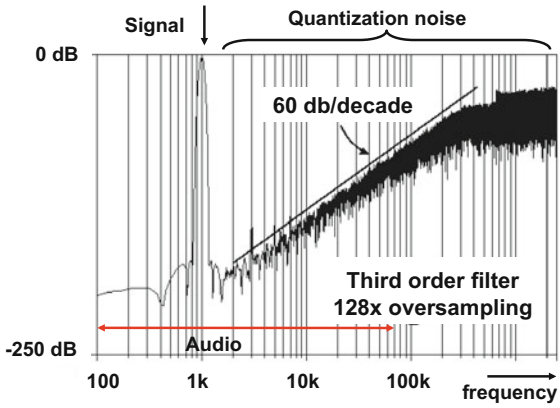


Fig. 10.16 Example of a frequency spectrum at the output of an audio sigma-delta modulator. The inverse third order filter characteristic can be observed in the noise pattern

If the amplification of the digital-to-analog conversion is set to $d = 1$, the substitution of $z \leftrightarrow e^{j\omega T_s}$ results in a frequency domain representation:

$$Y(\omega) = \frac{cG(\omega)}{1 + cG(\omega)} X(\omega) + \frac{1}{1 + cG(\omega)} N_Q(\omega) \tag{10.26}$$

At high values of $cG(\omega)$ the quantization errors are suppressed by this amount and in that frequency range the output signal equals the input signal. The slope of the noise spectrum is the inverse of the slope of the filter. In the frequency band where the filter does not produce much gain, the quantization power in the output signal increases, see Fig. 10.16. An increase of noise above the original noise spectral density of the quantization is visible close to the half of the sample rate. Here also patterns are observed that resemble frequency-modulation products (a sigma-delta modulator behaves here as a VCO or even an FM-modulator).

The previous description of the output signal $Y(\omega)$ as a sum of the processed input source and an additional noise source is often split in separate transfer functions for each: a signal transfer function (STF) and a noise transfer function (NTF):

$$\begin{aligned} Y(\omega) &= \mathrm{STF}(\omega) X(\omega) + \mathrm{NTF}(\omega) N_Q(\omega) \\ \mathrm{STF}(\omega) &= \frac{cG(\omega)}{1 + cG(\omega)} \\ \mathrm{NTF}(\omega) &= \frac{1}{1 + cG(\omega)} \end{aligned} \tag{10.27}$$

Figure 10.17 sketches both transfer functions for a simple sigma-delta converter. Often system designers like to see a flat STF: the signal passes through the sigma-delta modulator without experiencing filtering. However, at high frequencies where the active elements lose their gain, bumps and other irregularities may appear. Also the loop filter choice will impact the STF. The NTF is designed to suppress the noise at the band of interest. In that band it mimics the inverse filter function.

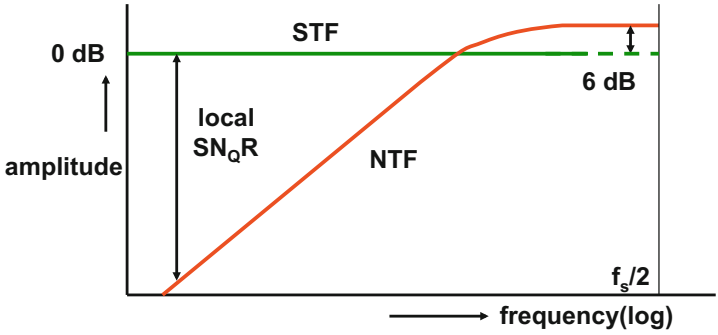


Fig. 10.17 The ideal signal transfer and a first order noise transfer function

Example 10.3. What indicates "instability" in a sigma-delta converter? What measures can be taken to recover from instability?

Solution. Instability for a sigma-delta converter is a non-oscillating output or an oscillating output whose time average does not follow the input signal. A sigma-delta converter is stable if it oscillates in a manner that the time average corresponds to the input signal.

For a (controlled) oscillation a loop gain of 1 and a phase delay of 180 is needed. In case of instability adjusting the state of the integrators is needed to return to a controlled oscillation. These adjustments can be implemented by limiting the integration or by a reset switch.

10.4 Time-Discrete Sigma-Delta Modulation

Since the first attempts to design oversampled converters, two main directions have evolved. The difference between the time-discrete and the time-continuous implementation is most notable in the filter design, see Fig. 10.18. The filter can be designed using time-discrete techniques such as switched capacitor, or time-continuous techniques as $g_m - C$ filters. It is obvious that a time-discrete system expects a sampled signal and that consequentially there is an anti-alias filter before the sampling circuit. In time-continuous modulators the loop filter can be used for anti-alias filtering. The definition of the poles and zeros of the filter with resistive and capacitive components is mostly an order of magnitude less accurate than a switched-capacitor filter based on the ratios of capacitors.

In this section sigma-delta modulators based on time-discrete filters are discussed, followed by time-continuous sigma-delta modulators in the next section. The discussion on the differences between the two is summarized in Table 10.3.

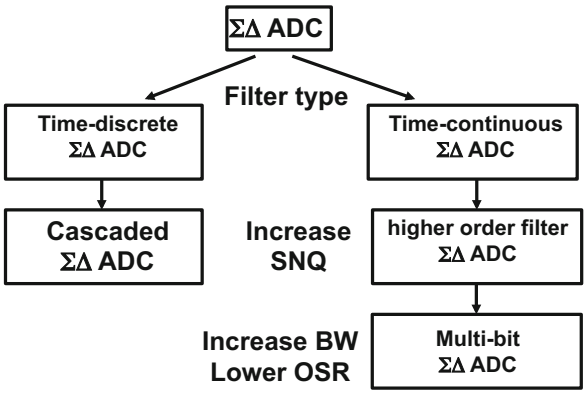


Fig. 10.18 A classification of sigma-delta conversion

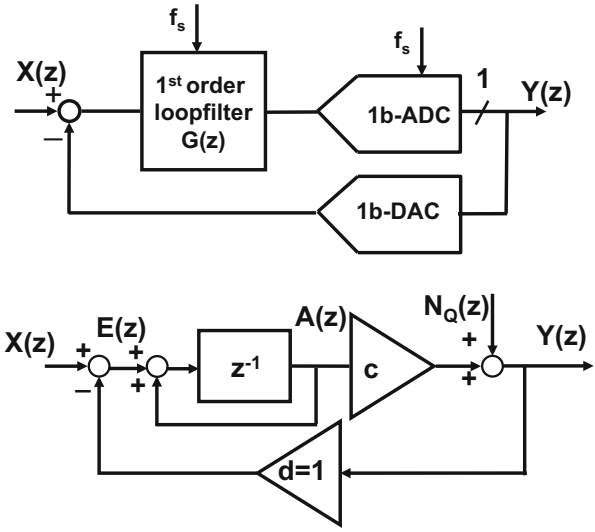


Fig. 10.19 A first order sigma-delta modulator and the equivalent analysis diagram

10.4.1 First Order Modulator

Figure 10.19 shows a first order time-discrete sigma-delta modulator and its linearized scheme. Comparison of the transfer function of the noise shaper and the sigma-delta modulator suggests that the filter function $G(z)$ can be obtained by rewriting $J(z)$. This is not the optimum strategy. A noise shaper is in fact not a feedback loop allowing more freedom when designing $J(z)$. The filter $G(z)$ in a sigma-delta modulator is the most essential part for obtaining a stable conversion.

The above transfer function, Eq. 10.24, is for the first order sigma-delta configuration split in the STF and the NTF. With $c \gg 1$ and $d = 1$ and an ideal integrator $G(z) = z^{-1}/(1 - z^{-1})$ these functions reduce to:

$$\mathrm{STF}(z) = \frac{Y(z)}{X(z)} = \frac{cG(z)}{1 + cdG(z)} \approx z^{-1}$$
$$\mathrm{NTF}(z) = \frac{Y(z)}{N_Q(z)} = \frac{1}{1 + cdG(z)} \approx 1 - z^{-1} \tag{10.28}$$

The signal passes unaltered through the modulator, delayed by one sample pulse. The noise is filtered, as can be seen in the frequency domain:

$$|\mathrm{NTF}(\omega)|^2 = (2\sin(\omega T_s/2))^2 = 2 - 2\cos(\omega T_s/2) \tag{10.29}$$

At half of the sample rate $\omega T_s/2 \approx \pi/2$, the NTF equals 2, see Fig. 10.17.

The total noise power in a band from 0 to f_b is found in the same way as for the noise shaper:

$$\int_{f=0}^{f_b} \frac{V_{LSB}^2}{6f_s} \left|1 - e^{-2j\pi f T_s}\right|^2 df \approx \frac{V_{LSB}^2}{12}\frac{\pi^2}{3}\left(\frac{2f_b}{f_s}\right)^3 = \frac{V_{LSB}^2}{12}\frac{\pi^2}{3\mathrm{OSR}^3} \tag{10.30}$$

The NTF is in this example a first order high-pass filter: suppression of noise close to DC while the noise is two-times amplified close to half of the sampling rate.

10.4.2 Second Order Modulator

A first order sigma-delta modulator oscillates stably as long as the input signal is sufficiently within the range of the feedback digital-to-analog converter. The addition of a second integrator stage, Fig. 10.20 with n = 2, creates a loop where the negative feedback (180°) plus the contribution of two integrator stages ($2 \times 90°$) equals 360°, see Fig. 10.21. It is clear that a small extra delay in the loop will push the modulator over 360°. The quantizer will take decisions on a signal that is too

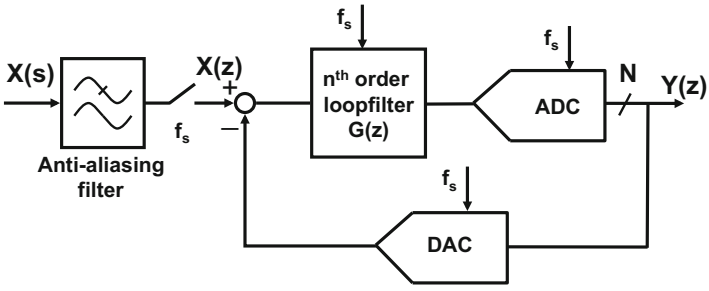


Fig. 10.20 Basic scheme of an *n*-th order loop filter sigma-delta modulator

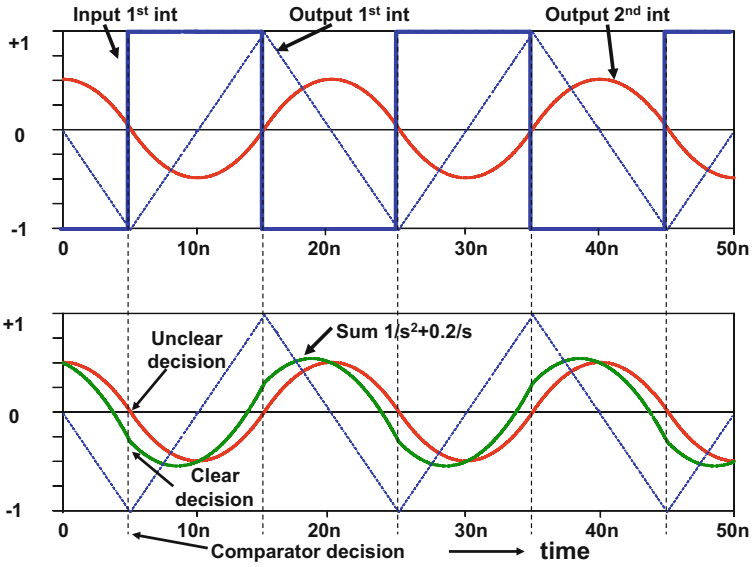


Fig. 10.21 *Upper*: the input to the first integrator is twice integrated and 180° in phase shifted. If this signal is fed into the quantizer (e.g., comparator), the zero-crossing can result in an arbitrary decision. *Lower*: a sum is formed of 1× the second integrator output plus 0.2× the output of the first integrator. The result is shifted some 15° backwards. The zero-crossing of the summed signals passes earlier through zero and will result in a correct quantization

late and create unwanted behavior. A second order modulator is only conditionally stable and proper conversion depends on input signal level, feedback, and delay.

In the lower half of Fig. 10.21 the input of the comparator consists of the output of the second integrator summed with 20 % of the output of the first integrator. This will create a total phase shift over the integrators and summing node of approximately 345°. And the predictable quantization (comparison) leads to a stable oscillation as in a first order modulator. Filter topologies for second order time-discrete sigma-delta modulation combine first order integrated signals to reduce the overall phase shift to below 360° and guarantee proper operation. In time-discrete filters the design of higher-order filters turning back to first order behavior requires zero's near $f_s/2$, which is cumbersome and can lead to unpleasant filter coefficients [312–314]. Therefore most time-discrete modulators use only first or second order topologies. Other converter topologies for higher-order sigma-delta modulation are discussed in the next section. In time-continuous sigma-delta modulation suitable higher filter-orders are easier to implement.

The noise transfer function for a simple second order time-discrete sigma-delta modulator for low frequencies is of the form:

$$\mathrm{NTF}(z) = \frac{Y(z)}{N_Q(z)} \propto (1 - z^{-1})^2$$

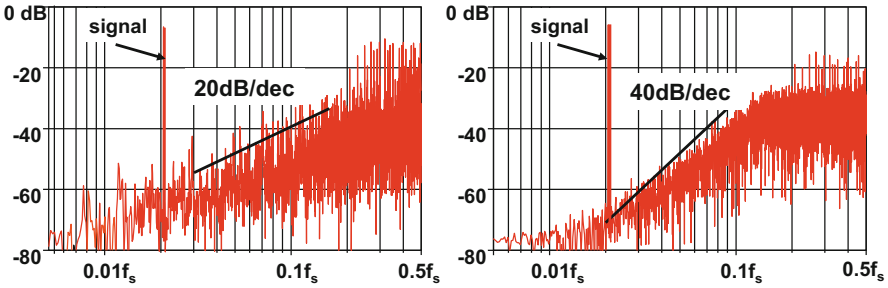


Fig. 10.22 The spectra of a first and second order sigma-delta modulator

The noise transfer function has now two zero's at DC ($z = 1$). The resulting noise power is approximated by:

$$\int_{f=0}^{f_b} \frac{V_{LSB}^2}{6f_s} |1 - e^{-2j\pi f T_s}|^4 df \approx \frac{V_{LSB}^2}{12} \frac{\pi^2}{5} \left(\frac{2f_b}{f_s}\right)^5 \tag{10.31}$$

which is a similar description as was found for a second order noise shaper, see Fig. 10.10.

The output spectrum of the second order sigma-delta modulator is compared to a first order modulator in Fig. 10.22. The first order spectrum still contains a lot of distinct frequency components pointing to a large correlation in the unwanted components. The second order spectrum shows a 40 dB/decade slope and much less idle tones. The dual integration of the signal and the two feedback paths from the quantizer create a much more complex pattern, which appears to contain less correlated products. Higher-order converters scramble the pattern even more due to the extra integration stages in the filters. Third and fourth order noise shaping shows therefore hardly any idle tones, see Fig. 10.16. Yet it is possible to devise signals where during short periods idle tones in second and third order modulators appear.

10.4.3 Cascade of Integrators in Feedback or Feed-Forward

In order to obtain a stable oscillation in a second order time-discrete sigma-delta modulator, the second order behavior must turn into a first order behavior near $f_s/2$. The simple cascade of two ideal integrators must be modified to a filter structure to get this behavior. In Sect. 2.7 two types of time-discrete filters were introduced. Finite impulse response filters are based on the addition of delayed copies of the signal and infinite impulse response filters use additional feedback of signals. In a similar way the second order filters inside a sigma-delta modulator can be

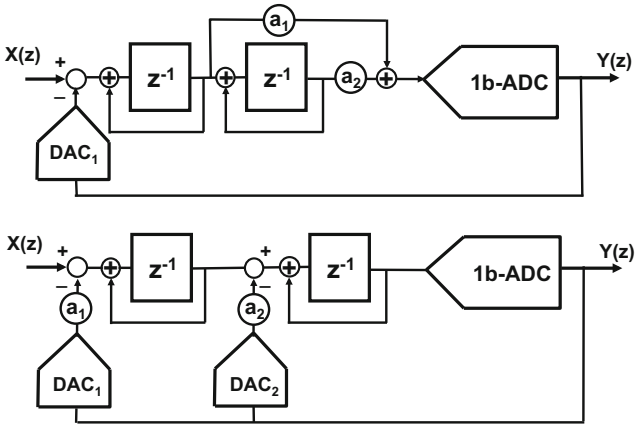


Fig. 10.23 Two methods to extend a first order sigma-delta modulator. *Upper*: the cascade of integrators feed-forward (CIFF) and *lower*: the cascade of integrators in feedback topology (CIFB)

classified: cascade of integrators in feed forward (CIFF) and cascade of integrators in feed back (CIFB). Figure 10.23 shows second order modulators constructed with CIFF and CIFB filters.

In the CIFF structure the input signal and feedback signal are subtracted in a summation node, Fig. 10.23 (upper). The filter function is obtained by creating a weighted summation (with coefficients a_1 and a_2) of the single and double integrated signals.

The input summation is in a practical design combined with the first integrator. This stage must be carefully designed as the sharp transitions of the 1-bit digital-to-analog converter easily can drive the opamp of the integrator into distortion. If the input stage provides sufficient amplification, the following stages process a filtered error signal. This signal is smaller and contains less high-frequencies. The noise and non-linearity of the following stages are less relevant due to the amplification of the first stage. This allows a low-bandwidth low-power design for these filter sections. In time-discrete filter realizations the switched-capacitor stages still must run at full sample speed. The advantage of low-bandwidth is therefore easier turned into low-power in a time-continuous filter as is discussed in the next section.

The filter transfer function for a second order CIFF filter is of the form:

$$G(z) = \frac{a_1 z^{-1}(1 - z^{-1}) + a_2 z^{-2}}{(1 - z^{-1})^2}$$

At low frequencies $z^{-1} \approx 1$ the a_1 term disappears to yield a second order shape, while for higher frequencies the a_1 term dominates and creates a stable first order behavior. The two poles at DC will lead to two zeros in the NTF, see Eq. 10.25. The topology of a sigma-delta modulator with a CIFF filter has a unity-gain feedback. The STF is therefore flat in the frequency range where the loop has sufficient gain.

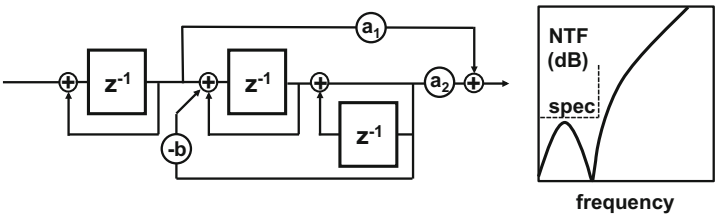


Fig. 10.24 A third order filter with feed-forward and two complex poles used to push the quantization power below the specification. The resulting NTF shows a double zero at the position of the complex filter pole

The zero in the STF transfer can lead to some peaking behavior in the STF transfer near $f_s/2$ leading to reduced signal ranges.

The filter function is shaped by the weighted addition of the outputs of the successive integrators. The additional summation point is loading the previous stages and requires careful design. The addition of internal signals can be extended if the input signal itself is fed into the summation point. This results in a specific feed-forward topology and is discussed in Sect. 10.8.3.

In advanced topologies there is a need to shape the transfer function beyond the DC point. If poles at other frequencies than DC are needed, a pair of complex poles must be used. The poles are created by modifying the ideal denominator:

$$\frac{1}{(1-z^{-1})^2} = \frac{1}{(1-2z^{-1}+z^{-2})} \Rightarrow \frac{1}{(1-(2-b)z^{-1}+z^{-2})} = \frac{1}{(p_1-z^{-1})(p_2-z^{-1})}$$

When $b > 0$ the roots of the denominator equation are at $p_{1,2} = 1 - 0.5b \pm j0.5\sqrt{4b-b^2}$, e.g., at ω_p where $cos(\omega_p) = 1 - b/2$. In the z domain all frequencies are normalized to $2\pi f_s$, so ω runs in the range $0, \ldots, \pi$ for the trajectory from DC to $f_s/2$. Implementation of this transfer function requires a local feedback path, see Fig. 10.24.

The filter function can also be realized by a feedback topology (CIFB), Fig. 10.23 (lower). Here two digital-to-analog converters are needed, one for every pole (pair) in the filter. The subtraction scheme allows more freedom to choose various signal levels. The feedback factor is not exactly unity, which affects the STF. Some peaking in the STF can occur.

In the CIFB topology each feedback path requires a summation node, designed for maximum performance. Moreover the subtraction of the feedback signal is distributed over several stages, that therefore have to operate at maximum linearity. Scaling of the signal is needed to optimally utilize the available signal range.

Figure 10.25 shows the feedback extension for a time-discrete second order topology. The feedback path coefficients are a_1 and a_2. If a_2 is too small, the first order content is insufficient to keep the modulator free of instability. If a_2 is too large, the modulator will behave as a first order system. The coefficients c_1 and c_2 are chosen to scale the signal after the summation point back into the range. With $c = c_1 = c_2 = 1$ the transfer is

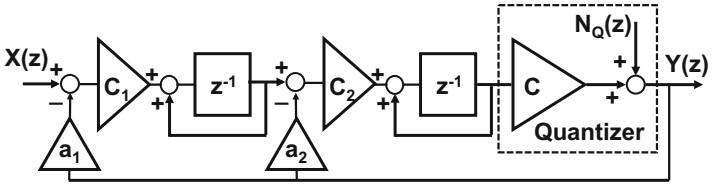


Fig. 10.25 Analysis diagram for a second order sigma-delta modulator with two feedback paths

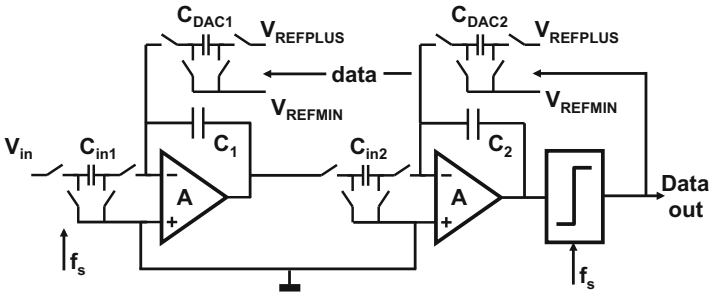


Fig. 10.26 A switched-capacitor realization of a second order sigma-delta modulator

$$Y(z) = \frac{X(z)z^{-2} + N_Q(z)(1-z^{-1})^2}{a_1 z^{-2} + a_2 z^{-1}(1-z^{-1}) + (1-z^{-1})^2}$$

Generally the filter formed by the feedback coefficients behaves as a low-pass filter for the input signal. An interesting choice is $a_1 = 1$ and $a_2 = 2$ resulting in a unity denominator.

For low frequencies the gain of the first integrator is high. The feedback path formed by the a_1 term plus integrator gain dominates over a_2. At higher frequencies the first integrator looses gain and the path via a_2 becomes important, which creates the first order path in the loop. As the comparator also provides gain, the gain of the filter can drop below unity close to $f_s/2$. The loop formed by the a_2 coefficient is the highest frequency feedback path. Between the unity-gain frequencies of the slower first integrator and the faster second integrator a frequency band exists where the noise of the second integrator is not insufficiently reduced by the first integrator gain. The input referred noise of the second integrator is now dominant. Therefore the second integrator needs to be designed at a sufficiently low noise level, at the expense of power.

Figure 10.26 shows a basic realization of a second order switched-capacitor sigma-delta modulator. The input network samples the input voltage and transfers this charge in integrator capacitor C_1. A following stage implements the second integrator. The output of the comparator determines whether a positive or negative charge packet is fed back to both integrators. In switched capacitor technique the capacitor ratios determine the scaling and the filter coefficients a_1 and a_2.

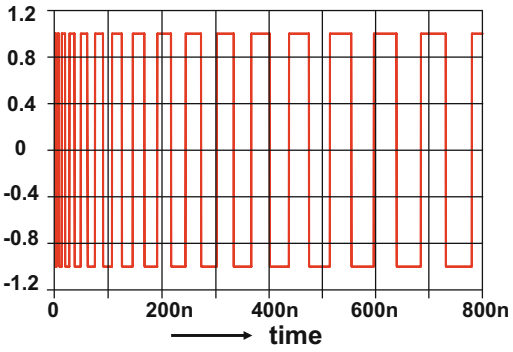


Fig. 10.27 The output signal of a second order sigma-delta modulator with a single feedback via two ideal integrators and zero volt input signal

Example 10.4. In a second order time-discrete sigma-delta modulator the first order feedback path is removed, so $a_2 = 0$ in Fig. 10.23 (lower) and the digital-to-analog converter causes a bit of additional delay in the loop. Draw the output signal.

Solution. Without a first order path the feedback signal will be delayed by two times 90°. Any additional delay will cause the feedback to be too late. So an edge generated by the comparator will be processed one clock period later. For zero-input signal a $f_s/2$ waveform is expected, but now successive edges will be processed with increasingly larger intermediate delays, see Fig. 10.27 for this typical form of sigma-delta instability.

10.4.4 Circuit Design Considerations

The analog summation point in the sigma-delta analog-to-digital converter must process high-frequency signal components at high linearity. The input summation point is often the most critical circuit block in the design. Issues are

- At the summation node the digital-to-analog converted signal is subtracted from the analog input. The combination of both signals can result in spikes that are more than double the input amplitude, due to the nature of the sampled and delayed signals.
- Bandwidths and signal swings of opamps or OTAs must match these linearity and noise requirements.
- The dynamic aspects have to be taken into account such as slewing of the currents.
- There is no gain yet at this node, so the impedance levels have to be designed to a thermal noise level that is acceptable for reaching the performance.

Designing the input with a sufficiently low-distortion requires to spend a lot of power.

Often some form of scaling is needed to reduce the signals to levels that can be handled by the integrators [315]. Another approach is to add gain before the quantizer and to use a scaled digital-to-analog conversion output signal. The scaling operation sacrifices useful dynamic range. The thermal noise contribution of the analog circuits becomes relatively more important with the risk of losing signal-to-noise ratio. Without scaling, some topologies (prominently CIFB structures) are limited by non-linearities and overload conditions. Architectures such as the feed-forward topology in Sect. 10.8.3 limit inherently the internal voltages and are less prone to non-linearity.

By far most filters in time-discrete sigma-delta modulators are designed with switched-capacitor techniques. As the filter switches at the oversample rate, the main issues are settling and power consumption. Traditional opamps or OTAs in switched-capacitor filters must deliver sufficient gain and current to optimally use this technique. Alternatives are inverter-based gain stages and combinations of active and passive filters: just capacitors and switches. Although the alternatives demonstrate good power efficiency [316], superior performance is still lacking.

In sigma-delta converters with very high sample rate, the comparator will be playing a more important role. A comparator, just as in Nyquist rate conversion, should generate an instantaneous decision. However, a delay period can arise or even a non-decision, as described in Sect. 8.1.3. The delay will cause phase-margin impairment in the loop and is in high sample-rate designs compensated by an excess-delay path, see Sect. 10.5.4.

Meta-stability in the comparator generates a noise-like component and can either be taken into the design as a parameter [317] or must be reduced along the same line as in Sect. 8.1.3.

10.4.5 Overload

In the early realizations of sigma-delta conversion the quantizer is a comparator. The output signal consists of a one-bit signal with two levels: either $0, 1$ or more symmetrical $-1, +1$. In Fig. 10.28 a sine wave is shown with such a one-bit

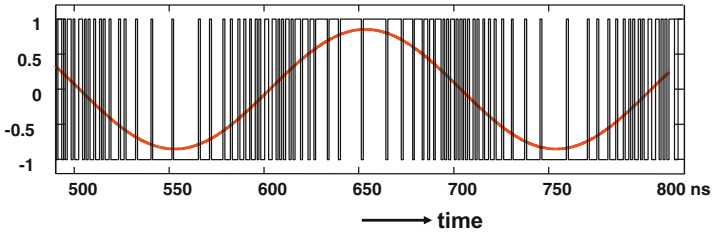


Fig. 10.28 A sine wave with an amplitude of 85 % of the full scale is represented by a pulse-density signal at the output of the feedback digital-to-analog converter

representation. The sine wave is here at 85 % of the range of the feedback digital-to-analog converter. It is clear that a sine wave that reaches the maximum levels of the digital-to-analog converter cannot be properly represented, as the modulator has no modulation freedom anymore at the maximum excursion of the signal. Another interpretation is that due to the fact that at the extreme levels many identical pulses are required, the effective oversampling frequency is strongly reduced, with unpleasant consequences.

The maximum peak–peak level of input signal with respect to the range of the feedback digital-to-analog converter that is correctly converted is characterized by the maximum overload level $\alpha_{OL} \leq 1$. The overload level is mostly expressed in dB below the maximum and ranges between -2 and -4 dB. On top of that there is another factor of 0.7 (-3 dB) between the power of a block wave and the power of an equal amplitude sine wave. The actual achievable overload level depends on the implementation of the converter, e.g., the filter order. Input levels that (temporarily) exceed the overload level will cause the internal signals to clip. The output signal cannot be switched in a sequence that corresponds to the input. This situation is called instability and often manifests itself as a fixed output value, or a slow oscillation. During overload the integrators that form the filter will be driven into a saturated state and will recover only slowly after the overload condition has been removed. A short period of overload may result in a much longer period of instability, see Fig. 10.29.

The stable situation in a sigma-delta converter is a controlled oscillation with a frequency close to half of the sample rate. Stability in a sigma-delta is defined as:

$$V_i(nT_s) \text{ is bounded by } \pm V_{DAC,max}, \forall n, \forall i \tag{10.32}$$

Every internal node V_i must be bounded at every time moment nT_s to the maximum voltages the feedback loop can provide.[8]

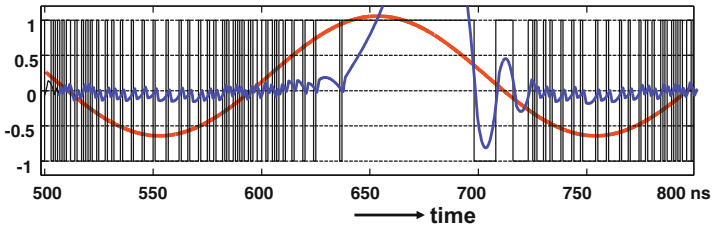


Fig. 10.29 A sine wave with an amplitude of 85 % of the full scale and 20 % bias creates an overload situation with instability. Shown are the input signal, the signal after the loop filter, and the digital code

[8]Normally a designer will try to use maximum signals throughout the entire circuit. In case scaled signals are used in some part of the circuit the value of V_i must be scaled as well.

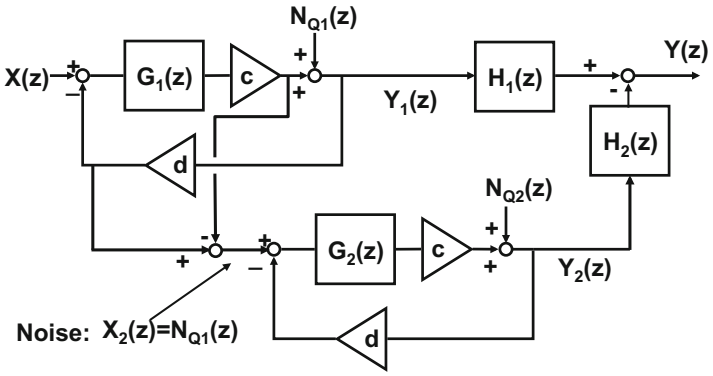


Fig. 10.30 The cascaded sigma-delta modulator or multi-stage noise shaper (MASH)

10.4.6 Cascaded Sigma-Delta Modulator

Extending the time-discrete sigma-delta converter to a third or higher order is possible; however, the loop filter has to be designed in such a way that a stable oscillation is guaranteed [314, 318, 319]. This can be done by inserting zeros in the transfer function. Yet, stability problems may occur. The gain of the quantizer in this analysis is assumed to be linear. In reality the quantizer gain is a non-linear function and varies with the input amplitude. An indication (but not more than that) for stability of 1-bit sigma-delta converters is based on simulations and called Lee's rule: $|\mathrm{NTF}(f)| < 1.5 - 2$ for all frequencies.

More noise power can be suppressed in a fixed bandwidth either by increasing the oversampling rate or by designing a higher-order loop filter turning back to first order. The cascaded sigma-delta approach [320] solves this issue in a third way: multi-stage noise shaping (MASH), see Fig. 10.30. Cascaded sigma-delta converters allow higher-order noise shaping without higher-order loop filters. A first sigma-delta converter uses a first or second order filter $G_1(z)$ in its quantization loop. The overall transfer of this sigma-delta modulator is given by Eq. 10.24. Although this sigma-delta loop suppresses the quantization errors $N_{Q1}(z)$ to a certain level, this error power can be further reduced by considering that this quantization error is a deterministic distortion signal that can be measured and subtracted from the output, see Fig. 10.30. The idea resembles a noise shaper, but where the noise shaper feeds the quantization power back, the cascaded sigma-delta digitizes the error and feeds the error signal to the output to cancel the error component in the signal.

The noise of the first sigma-delta converter is isolated by creating an analog copy of the output signal (via the digital-to-analog converter "d"). This signal is then subtracted from the input signal of the quantizer to yield: $X_2(z) = N_{Q1}(z)$. The quantization errors of the first modulator are now isolated in the analog domain and

are converted into the digital domain via a second sigma-delta modulator.[9] $Y_2(z)$ is the digital version of $N_{Q1}(z)$. Before this noise can be subtracted from the output signal $Y_1(z)$ the noise has to be shaped in frequency in a same way as $N_{Q1}(z)$ is shaped in by the loop of the first sigma-delta modulator. $Y_2(z)$ has to be divided by the loop filter characteristic $G_1(z)$ and $Y_1(z)$ has to be delayed to match the delay in the second modulator. The overall noise function is found as ($d = 1$):

$$\begin{aligned} Y_1(z) &= \frac{G_1(z)}{1+G_1(z)}X(z) + \frac{1}{1+G_1(z)}N_{Q1}(z) \\ Y_2(z) &= \frac{G_2(z)}{1+G_2(z)}N_{Q1}(z) + \frac{1}{1+G_2(z)}N_{Q2}(z) \\ Y(z) &= \frac{H_1(z)G_1(z)}{1+G_1(z)}X(z) + \left(\frac{H_1(z)}{1+G_1(z)} - \frac{H_2(z)G_2(z)}{1+G_2(z)}\right)N_{Q1}(z) - \frac{H_2(z)}{1+G_2(z)}N_{Q2}(z) \end{aligned} \tag{10.33}$$

With G_1, G_2 sufficiently large and $H_2(z)G_1(z) = H_1(z) = z^{-k}$ the conversion result is

$$Y(z) = X(z) + \frac{1}{G_1(z)G_2(z)}N_{Q2}(z) \tag{10.34}$$

The term z^{-k} represents the delay of k sample pulses in the second branch. The amount of noise suppression depends on the cancellation of the two paths that the noise of the first quantizer follows (the direct path and the path via the second quantizer and filter $H_2(z)$). If this cancellation is perfect, the remaining quantization errors originate from the second modulator. This noise can be reduced in the same way in a third loop. Yet, if a 1 % mismatch occurs in the cancellation, the first order noise will contribute on a level of 40 dB below the original first modulator performance (which still can be an acceptable performance). In [322] the effect of a matching error between the two filters is estimated as 0.8 dB loss in $\mathrm{SN}_Q\mathrm{R}$ for every 1 % mismatch. Figure 10.31 summarizes the signal flow in a cascaded sigma-delta modulator.

A cascaded converter with a second order filter in the main loop and a first order filter in the second is denoted as a "2–1 cascaded sigma-delta converter" or "2–1 MASH." The total noise transfer is equivalent to a third order noise shaper as in Fig. 10.10. The problem in the realization of analog-to-digital converters with cascaded sigma-delta modulation is therefore in the matching of the analog filter $G_1(z)$ with the digital counterpart $H_2(z)$. In switched-capacitor technique the matching of the analog and digital filter functions depends on the quality of the capacitor ratios and high-quality results can be reached [323].

[9] Variants exist where other analog-to-digital converters are used to digitize the quantization error [321], without much limitation.

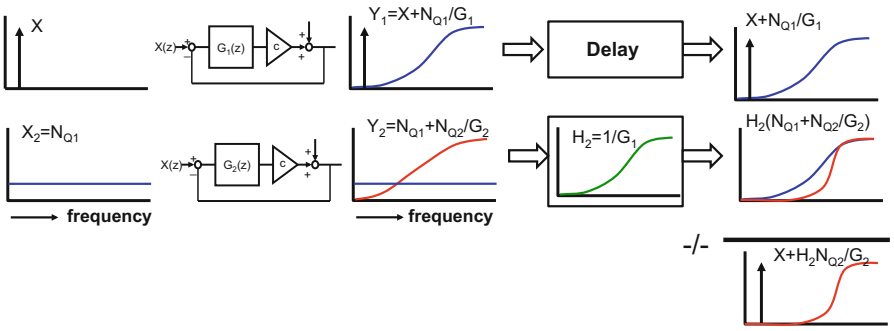


Fig. 10.31 Signal flow in a cascade sigma-delta modulator

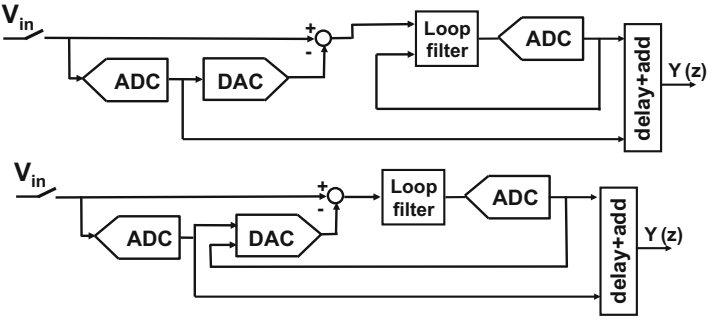


Fig. 10.32 *Upper*: a $0-L$ MASH converter [326], *below*: the zoom-ADC [327, 328]. Combining the digital streams requires addition, some delay and potentially overrange correction

A time-continuous cascaded sigma-delta converter needs a calibration mechanism to trim the digital filter to the analog filter [324] or careful design [325].

Figure 10.32 (upper) depicts a $0-L$ MASH cascaded sigma-delta converter. From a cascaded sigma-delta converter perspective, this topology has a zero-order filter in the first loop and an L-th order filter in the second loop [326]. From another point of view this is a subranging topology as in Sect. 8.3 with a sigma-delta converter as a second-stage converter. Or even as a noise shaping structure, where the quantization error of the first stage is digitized and subtracted. Either way, the idea is to reduce the input amplitude of the signal entering the sigma-delta modulator and ease the power-linearity trade-off of the second-loop filter design. Given the necessary sampling in the front-end, the filter is mostly a CIFB switched-capacitor topology. Also the matching between coarse and fine range, a standard problem in a subranging topology, must be properly addressed.

Gönen et al. [328] uses a third order feed-forward filter. Therefore in this "zoom analog-to-digital converter" the digital-to-analog converter in the first stage can be connected to the digital-to-analog converter in the sigma-delta feedback path. This design achieves an SINAD = 98 dB in 20 kHz bandwidth with an excellent F.o.M.

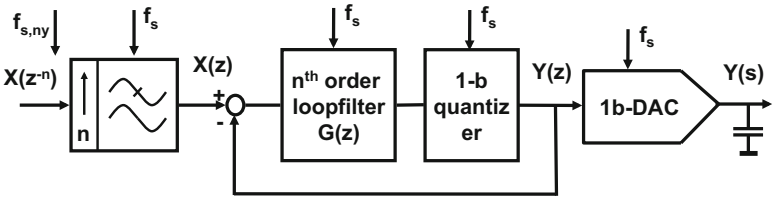


Fig. 10.33 A digital-to-analog sigma-delta modulator with single bit output

Example 10.5. Determine a suitable order of the analog filter and oversampling ratio to obtain 100 dB SNR gain in the desired bandwidth.

Solution. With the help of Fig. 10.10 various combinations are found: 16 times oversampling requires a fifth or sixth order, while 256 times oversampling needs only a second order filter.

10.4.7 Sigma-Delta Digital-to-Analog Conversion

In a digital-to-analog converter based on sigma-delta conversion a lot of elements return that were present in the analog-to-digital sigma-delta converter. The loop filter is now implemented in the digital domain, just as the quantizer. The main function of such a sigma-delta converter is the conversion from one digital word width at f_s to a smaller word width, e.g., 1-bit at OSR$\times f_s$, see Fig. 10.33. As no time-continuous input for high-frequency sampling is available, first the low sample-rate input signal must be up-converted to a higher sample rate, as described in Sect. 10.1. At the higher sample rate either noise shaping or sigma-delta modulation can be used to reduce the word width (quantizing), while keeping the spectral signal-to-noise ratio in a limited bandwidth in tact. The resulting bit-stream, e.g., a one-bit-stream, contains the required signal. In some applications with a high oversampling ratio, a passive filter is sufficient to obtain an analog signal of high quality. As mentioned before, the one-bit output signal contains a lot of unwanted energy and care has to be taken that this energy is correctly removed.

Various examples show the potential to reach the 16–18 ENOB level for audio bandwidths [329, 330] with different variations on Fig. 10.33. Kup et al. [329] uses a third order filter with complex poles, an OSR of 192 and a switched-capacitor output stage and achieves a 102 dB SINAD in a 2.5 µm BiCMOS process.

The 1-bit output format removes the need for accurately spaced conversion levels. Yet, handling the output bit-stream without distortion and providing a low-jitter restoration clock may present a challenge. A full 20-bit signal in a 20 kHz bandwidth would require a clock jitter of less than 8 ps_{rms}, which is realizable for a consumer equipment price level.

Another interesting option to create an analog signal while filtering off the high-frequency components is the semi-digital converter in Sect. 7.3.7. The introduction

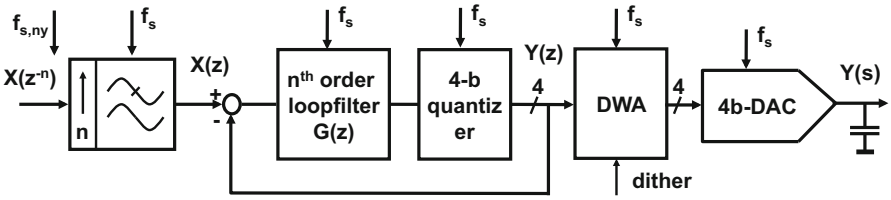


Fig. 10.34 A digital-to-analog sigma-delta modulator using a multi-bit output linearized with a data-weighted averaging algorithm, dither or scrambling [331, 332]

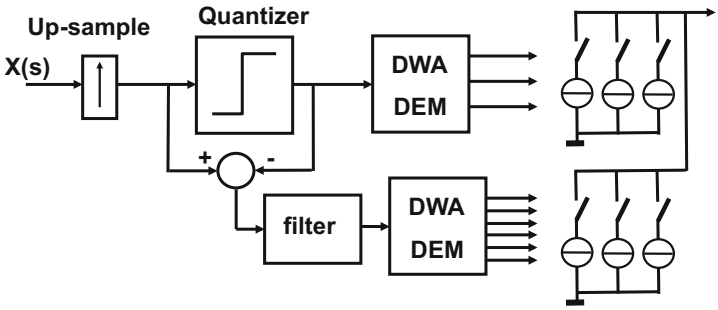


Fig. 10.35 Cascading is applied in digital-to-analog conversion

of data-weighted averaging, see Sect. 7.6.3, has allowed to use a multi-bit output, thereby reducing jitter sensitivity and pushing the dynamic range over 100 dB [331, 332], Fig. 10.34.

10.4.8 Cascaded Digital-to-Analog Conversion

The cascaded analog-to-digital converters in the previous sections recombine their signals in the digital domain. In digital-to-analog conversion a similar technique is possible, where current steering converters are used to combine the two paths. Addition of currents is a matter of connecting current sources together. Figure 10.35 shows a potential set-up. After up-sampling the data stream is split into an MSB part and an LSB part. Both paths use sigma-delta modulation to create high-speed low-resolution data streams. The MSB data stream serves to create a low-distortion path using only a limited number of (calibrated, DWA or DEM) current sources. In [161] the MSB stream is reduced to a single bit. The quantization error is processed in filters to remove the unwanted noise contributions and is also converted into currents. At the summation node the LSB stream now compensates the quantization errors in the MSB stream. An interesting example is found in [161].

10.4.9 Decimation

Similar to the situation in time-division digital-to-analog conversion in Sect. 7.5 the one-bit representation of a sine wave generated by sigma-delta modulation shows a lot of quantization or unwanted products. Comparing the power consumed in a resistor R of a digital sequence switching between $+V_a$ and $-V_a$ to the power of an equal amplitude sine wave:

$$\text{Power of one-bit digital} \leftrightarrow \text{Power of maximum sine wave}$$

$$P = \frac{V^2}{R} = \frac{(\pm V_a)^2}{R} = \frac{V_a^2}{R} \leftrightarrow P = \int \frac{(\alpha_{OL} V_a \sin(\omega t))^2}{R} dt = \frac{\alpha_{OL}^2}{2} \frac{V_a^2}{R} \tag{10.35}$$

shows that more than half of the power in a one-bit signal consists of unwanted distortion. Next to this unwanted power, the overload factor $\alpha_{OL} \approx 0.7$ causes another -2 to -4 dB loss in the practical implementation.

The one-bit representation of a sigma-delta output signal is therefore largely composed of high-order distortion products. Any non-linearity in the feedback of the converter can easily create intermodulation between these higher-order sampled distortions. Spurs will appear in the signal path. Therefore the quality of the digital-to-analog converter, especially concerning distortion, is crucial in the design of high-performance sigma-delta modulators. Also the digital processing of such a signal in the succeeding signal path requires attention for all non-linear operations such as rounding.

The large amount of quantization noise located at high frequencies must be removed before the signal can be processed. Moreover, many signal-processing applications are optimized for PCM-type digital representations as a Nyquist converter will deliver.

In order to remove the high-frequency quantization-noise and return to a PCM-type sample format, the high-speed data from the modulator's out has to pass through a decimation filter. The fundamentals of the decimation process have been eluded in Sect. 2.3.3. The digital filter adds differently weighted low-resolution samples together. After filtering has taken place, Fig. 10.36, data is discarded by selectively removing samples (decimation). The filter is designed to separate the signal bandwidth from the high-frequency quantization components. In low-sample-rate sigma-delta modulation, this filter can be designed with the available CAD tools, e.g., as an FIR filter. During the filter operation a weighted sum of a (large) number of samples is formed. When this process runs at the speed of the modulator, the digital power consumption will be comparable or even exceed the analog power.

However, today's GigaHertz-rate advanced sigma-delta modulators would pose severe speed problems on filter realizations in standard digital libraries. The Hogenauer or cascaded-integration-comb (CIC) filter is in those cases an efficient solution to the decimation problem, Fig. 10.37. In the high-frequency domain the signal is integrated, causing a low-pass characteristic. After decimation the

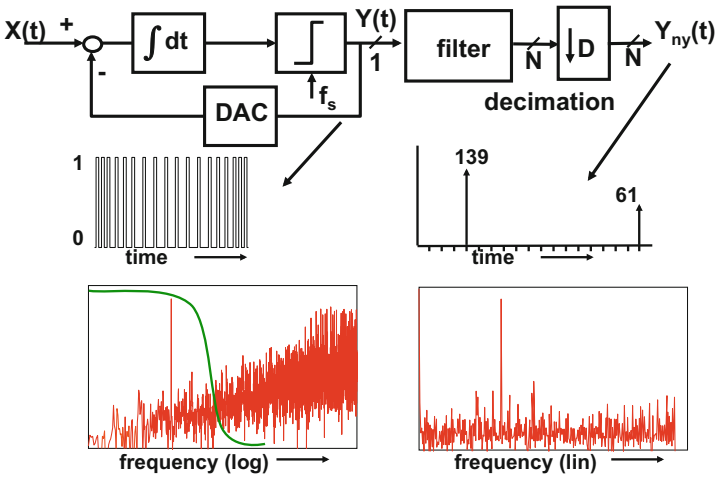


Fig. 10.36 Decimation after the modulator allows to transform the one-bit data stream in PCM-type samples. On the *lower left side* is the time and frequency representation of the one-bit data stream. On the *right side* the filter and decimation stage have created a Nyquist-rate sample sequence

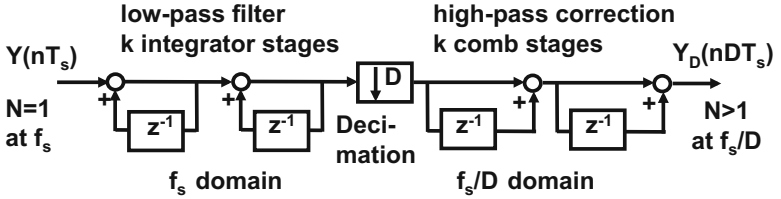


Fig. 10.37 Decimation with a Hogenauer structure

comb filter corrects the frequency behavior in the desired band. Obviously only summation/subtraction and storage functions are need, no power hungry multipliers for FIR-coefficients. Of course the integrators need huge word widths in order to accumulate the signal over a longer time period, and some means has to be found to avoid overflow. The transfer function of this filter is

$$|H_{CIC}(f) = \left| \frac{\sin(\pi D f T_s)}{D \sin(\pi f T_s)} \right|^k \tag{10.36}$$

An interesting solution that avoids the complexity of the FIR structure, and does not need the corrections of the integrator-comb structure, is the half-band filter cascade in Fig. 10.38. A half-band filter of Sect. 2.7.2 cleans the frequency range between $f_s/4$ and $3f_s/4$ of unwanted signal components such as quantization noise. After that the sampling rate can be reduced by subsampling from f_s to $f_s/2$. Now the same procedure is repeated: a simple half-band filter eliminates the unwanted

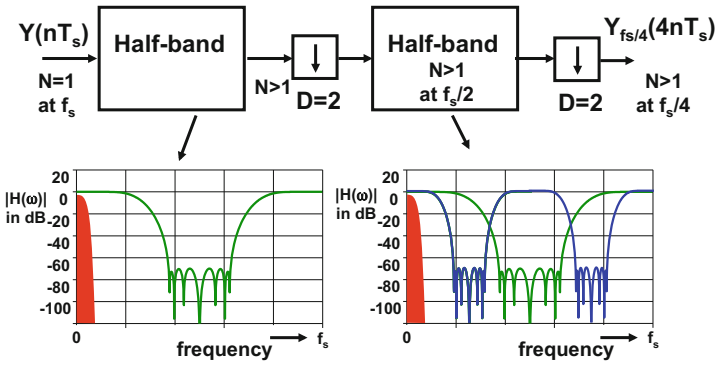


Fig. 10.38 Decimation with a cascade of half-band filters

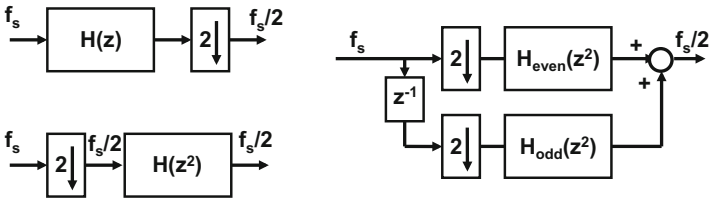


Fig. 10.39 Subsampling can be realized by selecting every second sample (*upper-left*). Sample selection and decimation can be interchanged (*lower-left*). More advanced methods split the filter in an even-sample and odd-sample portion

components and the rate is halved. A cascade of various subsample sections achieves a large overall subsample factor.

High-speed subsampling can be achieved via half-band filtering at full-speed and selecting every second sample, see Fig. 10.39 (upper-left). This method is not efficient, half of the calculated output is unused. In filter theory the "noble-identities" allow to reverse the order of filtering and subsampling and create an energy efficient solution to the problem of implementing energy-efficient subsampling, see Fig. 10.39 (lower-left) [17]. An efficient alternative is to split the filter into two sections for the odd and even samples, Fig. 10.39 (right).

Example 10.6. A fast N-bit analog-to-digital converter runs at f_s. The succeeding logic cannot handle this speed and processes one in four samples. What will this do to the SN_QR and the spectral power density of the quantization power? Now every input to the logic is created by an adder that sums four samples. Does this help? Due to the addition the digital word width grows by 2 bit to $N + 2$. Can the last bit be discarded?

Solution. The original converter will yield a quantization power $V_{LSB}^2/12$ in a bandwidth of $f_s/2$ and a spectral power density of $V_{LSB}^2/6f_s$. By using only one-in-four samples the quantization energy of the converter does not change, so the

SN_QR remains the same. The sample rate and thus the bandwidth will reduce by four. As a consequence the spectral power density will go to $4 \times V_{LSB}^2/6f_s$.

Adding up four consecutive samples means that the signal amplitude increases by four and its power by 16, while the (uncorrelated) quantization errors add up root-mean-square giving a four times higher quantization power and a four times higher spectral power density which now goes to $16 \times V_{LSB}^2/6f_s$. In first order the SN_QR improves by a factor four over the single selected sample or 1 ENOB. The signal quality is improved by 1 bit at the expense of a four-times lower sample rate and Nyquist bandwidth.

The digital word width increases by 2 bits. However, removing the lowest bit will quantize the signal and add some more quantization power.

10.5 Time-Continuous Sigma-Delta Modulation

10.5.1 First Order Modulator

Time-discrete sigma-delta modulation assumes a sampled data stream at its input. All internal processes are synchronized to the sample rate. The alias filtering necessary for sampled data systems must be performed before the data enters the sigma-delta modulator. In contrast, time-continuous sigma-delta modulation uses a loop filter based on time-continuous components. The sampling takes place after the filter, see Fig. 10.40. Mostly the sampling is implemented by the decision making of the quantizer. The conversion system is now partly time-continuous and partly time-discrete. Figure 10.41 shows the wave forms for a zero-input signal. In contrast to the synchronous mode of working in a time-discrete sigma-delta converter, here the timing in the loop is only synchronized by the quantizer. The quantizer sends out a digital signal that is converted to analog levels $D(s)$ by the digital-to-analog converter with minimum loss of time. This signal gets inverted at the summing node resulting in $E(s)$ and is integrated to a triangular shape $A(s)$. At the sample moments the comparator decides on the polarity resulting in the time-discrete signal $Y(z)$. The frequency of the round going signal is $f_s/2$. The fundamental signal before and after the comparator $A(s) \rightarrow Y(z) \rightarrow D(s)$ is 90° shifted due to the conversion of

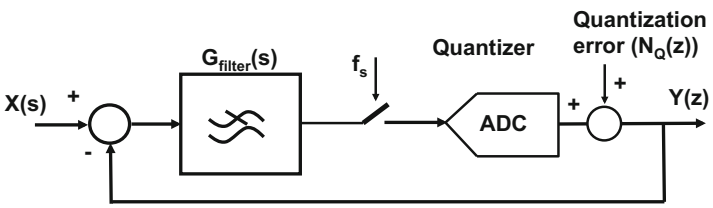


Fig. 10.40 A time-continuous sigma-delta modulator samples after the filter

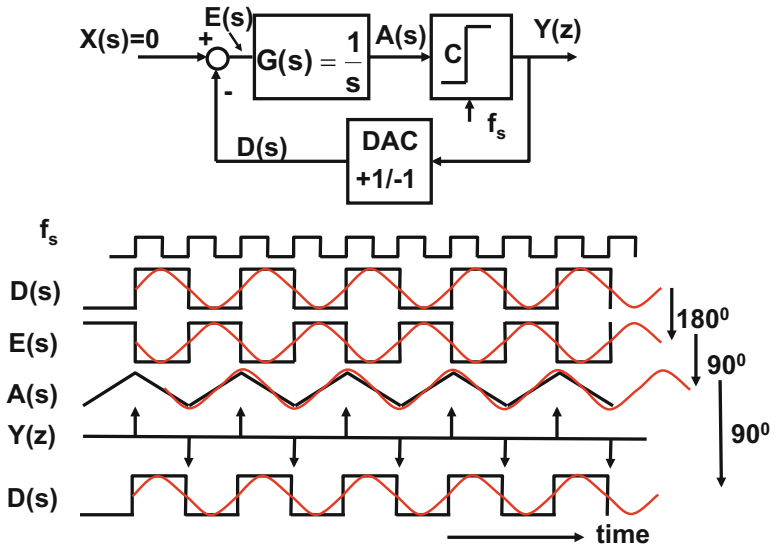


Fig. 10.41 The signal flow in a first order time-continuous sigma-delta modulator. The quantizer switches on the rising edge of f_s. The fundamental frequency for all *block* and *triangle shaped* signals is indicated to see the phase relations. For clarity the $D(s)$ signal is repeated

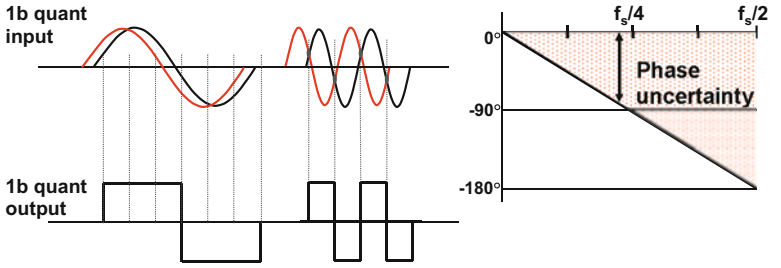


Fig. 10.42 On the left-hand side a slow signal (*red*) and a slightly time or phase shifted copy (*black*) both result in the same comparator decision. The same time delay allows a much larger phase shift for a higher frequency signal (*middle*), as $\phi = 2\pi f T_s$. The time delay before sampling f_s translates in an increasing phase tolerance for each increasing signal frequencies (*right*)

the triangular shape to a block wave. Part of the phase shift to reach 360° for an oscillating loop is accomplished through the waiting time before the sample pulse activates the quantizer. Therefore some phase margin or waiting time exists after the signal passes the filter. The comparator decides on the edge of the sample pulse. With respect to this sample pulse signals will be early in a first order modulator, see Fig. 10.42. This margin can be anything between zero and the period of one sample pulse. The margin is often referred to as "phase-uncertainty," which is a misleading term as there is no uncertainty involved.

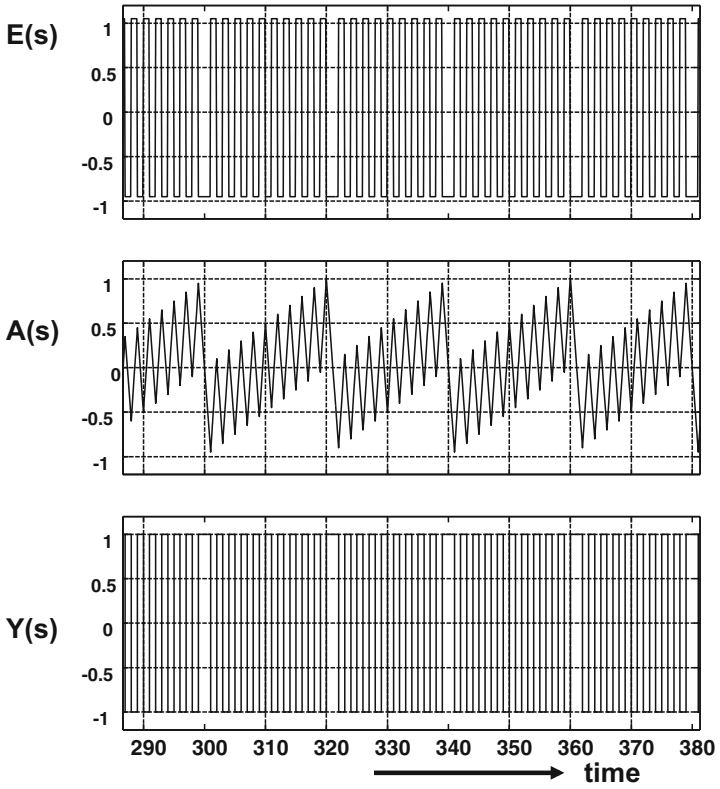


Fig. 10.43 The signals in a first order time-continuous sigma-delta modulator with a small DC input signal

The waiting time in a first order modulator is partly consumed by the delay of the integrator. This implies that the delay of a second integrator in the loop would just fit. Any additional delay will corrupt the stability.

In the presence of some DC-offset at the input as in Fig. 10.43, the wave forms in the time-continuous sigma-delta converter resemble the wave forms in the time-discrete case in Fig. 10.13. The integration of the high-frequency component and the input signal is now well visible.

The STF and a NTF are formulated in the s or Laplace domain:

$$\begin{aligned} Y(s) &= \mathrm{STF}(s)X(s) + \mathrm{NTF}(s)N_Q(s) \\ \mathrm{STF}(s) &= \frac{cG(s)}{1 + cG(s)} \\ \mathrm{NTF}(s) &= \frac{1}{1 + cG(s)} \end{aligned} \tag{10.37}$$

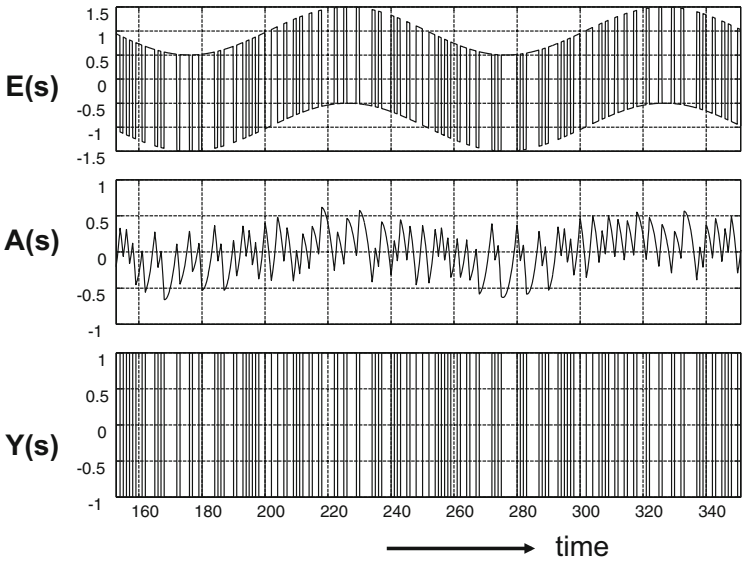


Fig. 10.44 The signals after the input summation point, after the second order filter and at the output of the 1-bit quantizer for a second order sigma-delta modulator

where c is the gain of the comparator, as in Fig. 10.15. The complication of this description is the accuracy of the modeling of the sampling and quantization after the time-continuous filter by an addition of noise and a gain factor.

A second order time-continuous sigma-delta modulator is created by adding an integration stage in a feed-forward manner, see Fig. 10.23 (upper). Figure 10.44 shows the internal signals for a sine wave input. The input summation node has to process both the analog input signal and the feedback 1-bit signal without distortion. The second order filter will suppress the quantization errors better. This leads to less correlated error signals (less sine wave visible) as can be seen from Fig. 10.45.

The time-continuous loop filter has to meet the specifications within the desired bandwidth, which is a factor $2 \times OSR$ lower than the sample rate. The design of a time-continuous filter can be derived from the time-discrete implementation [333] for first and second order filters. Table 10.2 shows the equivalence for first and second order integration.

In the time-discrete domain, where the filter is built in switched-capacitor technique, the switches run at the sample rate and the opamps need to settle at that speed. This requirement normally leads to a higher power consumption in the opamps and is less present in time-continuous filters. The configuration of a time-continuous sigma-delta modulator has an advantage that (a part of) the alias and blocker filtering is performed by the loop filter. Now noise and distortion mainly determine the power consumption of the opamp(s). The disadvantage of the time-continuous filter is that although its shape or the relative position of poles and zeros can be well controlled, but not the absolute position in the frequency band.

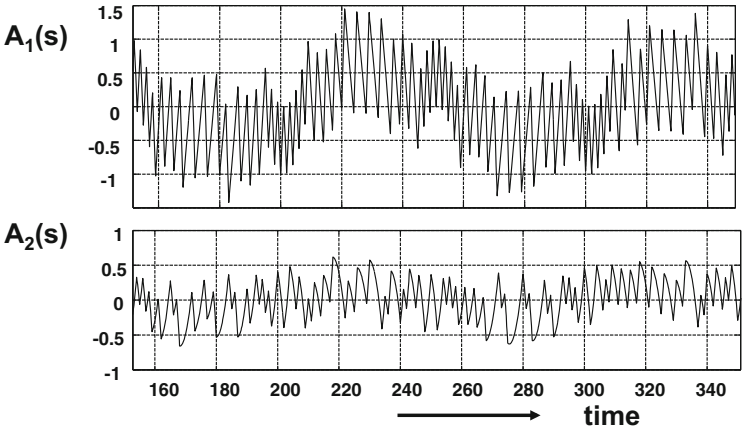


Fig. 10.45 A comparison of the signal after the filter in a first order and a second order sigma-delta modulator

Table 10.2 Equivalence between time-discrete and time-continuous integrators, from [333]

Time-discrete z-domain		Time-continuous s domain
$\frac{z^{-1}}{1-z^{-1}}$	$\Leftrightarrow$	$\frac{1}{sT_s}$
$\frac{z^{-2}}{(1-z^{-1})^2}$	$\Leftrightarrow$	$\frac{1}{s^2T_s^2} - \frac{1}{2sT_s}$
$\frac{z^{-2}}{(1-z^{-1})^2} + \frac{z^{-1}}{2(1-z^{-1})}$	$\Leftrightarrow$	$\frac{1}{s^2T_s^2}$

Note that these equivalences hold in the pass-band up to $f_s/2$ and also the differences due to T&H performance must be taken into account

A tolerance of 10 % in modern processes has to be taken into account and translates in some guard-banding in the specification of the bandwidth of interest. As the filter has fixed frequencies for its poles and zeros, the circuit will work only around one sample frequency. In a switched-capacitor filter the filter curve nicely follows the sample rate.

10.5.2 Higher-Order Sigma-Delta Converters

A prerequisite for the application of sigma-delta conversion is a sufficiently high oversampling factor of f_s over the signal bandwidth. The sigma-delta modulator does not reduce the quantization power as a whole, it only moves undesired quantization power from one frequency band into another band. For achieving low-noise levels in wider bandwidths without increasing the sample rate into the Gs/s range, in time-continuous sigma-delta converters the filter order must be increased.

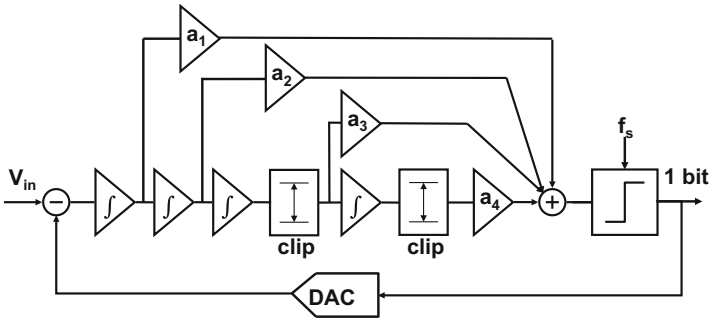


Fig. 10.46 A fourth order feed-forward sigma-delta modulator [334]

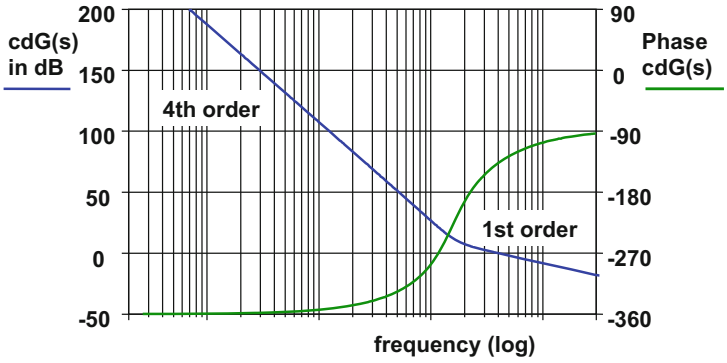


Fig. 10.47 The fourth order filter transfer function turns back to a first order transfer function close to the frequency region where the modulator acts as an oscillator

Figure 10.46 shows a sigma-delta converter with a fourth order feed-forward filter. The loop filter consists of a cascade of four integrators whose outputs are summed via coefficients. The comparator decides on the polarity of the output of the filter.

The design of the filter function has to take into account that the requirement for stable oscillation in the loop requires a feedback path with a total phase shift before the quantizer of less than 180° at the oscillation frequency (close to $f_s/2$). Therefore, the design of higher-order filters is possible if the filter is designed in such a way that it shows a high-order behavior for low frequencies, while turning back to a first order response near the passing of the 0-dB loop gain. In Fig. 10.47 the low-frequency noise shaping part of the filter is fourth order. The filter is designed to turn back to a first order and 90° phase at the frequencies where the 0 dB gain is reached. At frequencies close to $f_s/2$ a 360° overall phase shift is reached: 180° from the inversion at the summation node, $>90°$ from the filter, and the rest is consumed in the delay before the sampling moment, the phase uncertainty.

In case of instability due to an excess voltage at the input, the outputs of the last two integrators will saturate first and longest. The clippers in Fig. 10.46 effectively

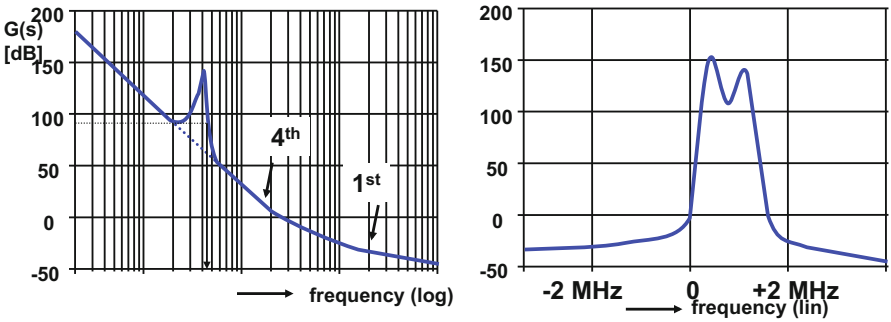


Fig. 10.48 Resonators and other filter elements are used to modify the loop filter curve. On the *left side* a pole in the filter increases the gain in a limited frequency range, thereby stretching the noise suppression. *Right*: a complex filter allows to create suppression in the positive frequency range. Example taken from [324]

switch off the contribution of these integrators to the overall filter function. The modulator will behave as a third or second order sigma-delta converter. Although the quantization suppression is temporary degraded, still some signal conversion is performed. This behavior is called "graceful degradation." After removal of the excess signal the integrator will recover and help again to shape the quantization errors.

The shape of the filters in a time-continuous sigma-delta converter depends on the ratios of the frequency determining elements. The absolute values of these components determine the pole and zero frequencies.

Many techniques known from analog filter design have been applied to modify the noise transfer characteristics. Inserting resonators allows to stretch the noise-suppression frequency range in Fig. 10.48 (left). More poles in the filter on different frequencies can be applied to shape the noise transfer. Poles in the loop filter turn into zeros in the NTF.

Quadrature signals are very common in advanced communication. Complex sigma-delta modulators, Sect. 10.8.1, use two parallel modulators driven by in-phase and quadrature signals. The filters of these two modulators are coupled to perform complex filtering. Figure 10.48 (right) shows a filter characteristic that suppresses only for positive frequency ranges [324]. Figure 10.49 shows the output simulation of a 64-times oversampled modulator with two zeros in order to stretch the useful bandwidth. Note that the filter turns back to first order behavior a decade in frequency before arriving at $f_s/2$.

The combination of a time-continuous filter and a time-discrete quantization creates a few additional problems in the Laplace analysis. Not only the definition of the gain factors c and d becomes more complicated, but also the time relation between the comparator's input and output is less trivial. The gain in the loop is relevant for the loop behavior. As the comparator interfaces between the analog and digital domain and is highly non-linear, the description of the gain is not a simple mathematical expression.

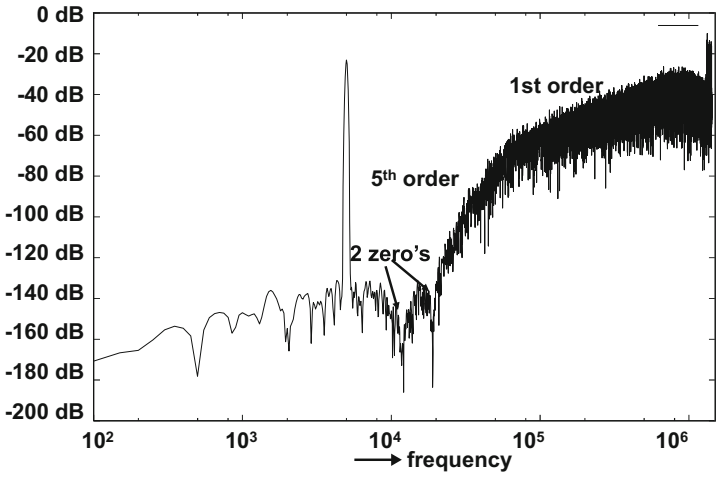


Fig. 10.49 The simulation of a 64-times oversampled spectrum with a fifth order filter and two zeros to stretch the low quantization bandwidth to 20 kHz. Courtesy: Breems/ v. Veldhoven

The comparator signals have no linear or linearizable relation, the gain must be determined by comparing the power functions. If cd incorporates the comparator and the gain of the digital-to-analog converter, where $T \gg T_s$, the power gain is

$$(cd)^2 = \frac{\int v^2_{quantizer,out} dt}{\int v^2_{quantizer,in} dt} = \frac{(\pm 1)^2}{(1/T) \int\limits_T v^2_{quantizer,in} dt} \tag{10.38}$$

It can be necessary to do a numerical comparison of the input power with the rather simple output power of the comparator/digital-to-analog converter. In a simulation example the gain does not vary much over the input amplitude range, see Figs. 10.50 and 10.51. Moreover the feedback loop concept tolerates some inaccuracy in this parameter. Another way of looking at this problem is that the comparator and digital-to-analog converter produce constant amplitude signals with a high content of $f_s/2$ related frequencies. This signal (plus the input) is fed into the filter and will produce a more or less deterministic input amplitude to the comparator. The rather small deviation in gain can be dealt with by requiring a gain guard band in the stability analysis.

Figure 10.52 shows the resulting design of a fourth order loop filter, including some guard banding for the gain inaccuracy. Also the phase uncertainty region is indicated. This region indicates where the phase must be in order to be in-time for sampling. The oscillation will chose a frequency where the overall gain is 1 and a phase of 180° is available. This is a necessary but not sufficient condition

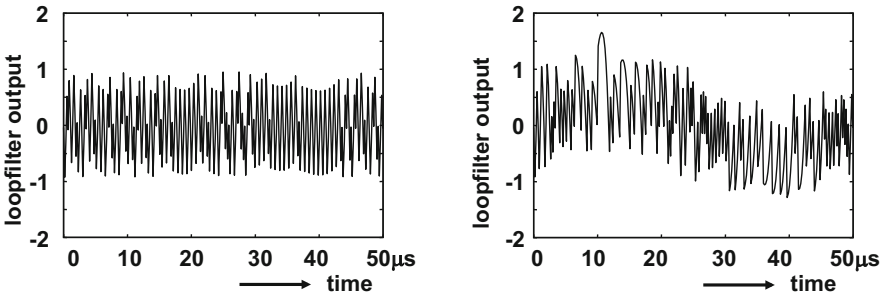


Fig. 10.50 The simulated input for the quantizer, *left:* with 0.1 % signal input to the modulator, *right*: with maximum input signal, from [334]

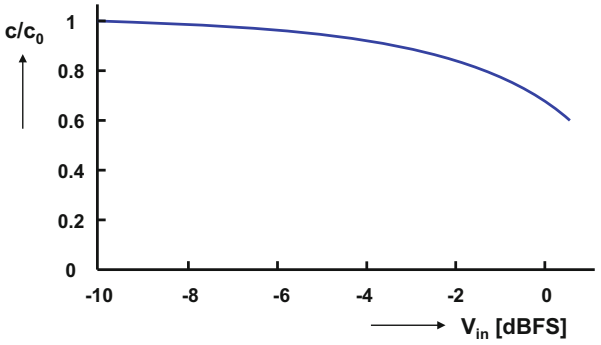


Fig. 10.51 The gain as function of the input signal to the modulator, from [334]

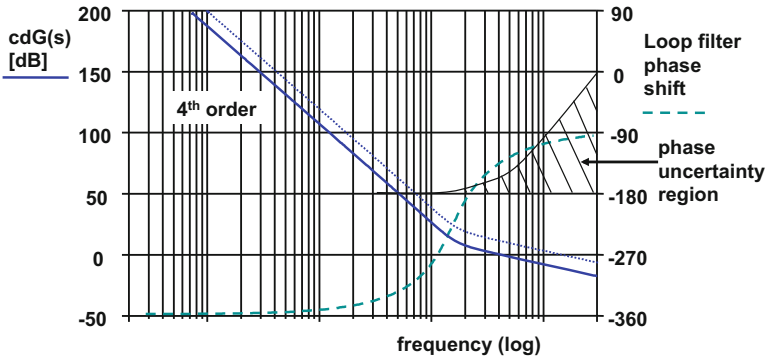


Fig. 10.52 The fourth order filter transfer function with gain tolerance and the phase plot including phase uncertainty

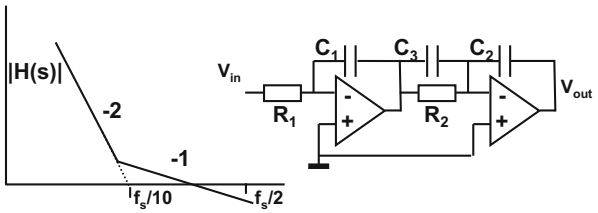


Fig. 10.53 Filter and transfer function for a second order time-continuous sigma-delta converter

for the loop to oscillate. Applying common practices such as Lee's rule does not eliminate the need to rigorously simulation of any non-linear circuit to verify proper functionality for all input levels.

In time-continuous sigma-delta modulation also the idea of cascading two or more loops has been explored [324, 336]. However, no obvious advantage has been demonstrated.

Example 10.7. A second order sigma-delta modulator runs at a sampling rate of 5 Ms/s. A signal band of 20 kHz is required. The comparator and digital-to-analog converter can be assumed ideal. The gain uncertainty and the sampling uncertainty of the comparator need to be taken into account. Set up a second order filter using capacitors, resistors, and ideal opamps, that gives a maximum SNR ratio in this band.

Solution. The oversample ratio is 2.5 MHz/20 kHz $=$ 125. With an ideal 2-order 1-bit sigma-delta modulator this would give 6 dB from the 1-bit quantizer and 92 dB shaping improvement, see Fig. 10.10. A 2-order integrator can be designed as a cascade of two opamp-R-C sections, Fig. 10.53. However, the loop would be unstable. Therefore an additional capacitor C_3 is added. So a component choice $R_1 = R_2 = 100\,\text{k}\Omega$ and $C_1 = C_2 = 3\,\text{pF}$ gives a unity gain for the two integrators of 500 kHz. $C_3 = 4$ pF will create a zero factor 3–4 before the unity-gain frequency. At $f_s/2$ the overall filter gain will be around -20 dB. The gain of the comparator has to compensate this attenuation. The filter will amplify 55 dB at 20 kHz, which is a loss of about 30 dB compared to the ideal situation. Figure 10.54 shows a simulation with this filter in a sigma-delta loop.

10.5.3 Digital-to-Analog Conversion in the Loop

A major disadvantage of time-continuous conversion relates to the digital-to-analog converter. In a time-discrete switched-capacitor implementation the feedback charge is defined by the capacitor and the reference voltage. The timing is of secondary importance as the charge will anyhow be transferred into the summation point. In a time-continuous converter the feedback from the digital-to-analog converter

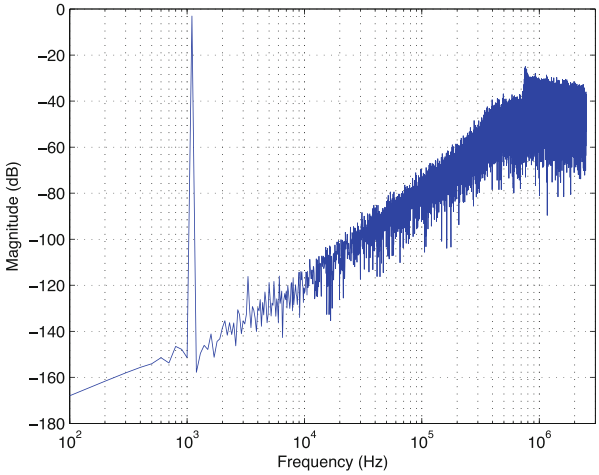


Fig. 10.54 Conversion of a 1.1 kHz input signal. Courtesy: Hugo Westerveld U Twente

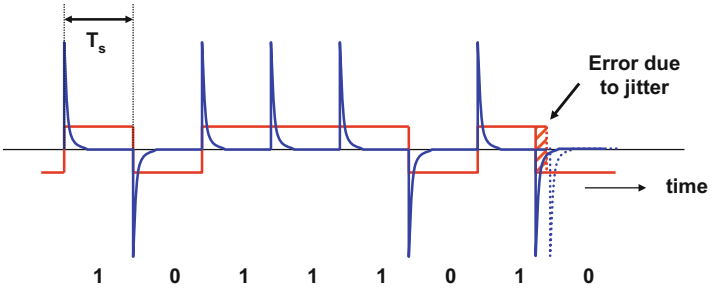


Fig. 10.55 A time-continuous DAC feedback signal is a continuous pulse. A switched-capacitor DAC generates charge spikes. Jitter just shifts the position of a switched-capacitor pulse, while it modulates the volume of a time-continuous pulse

is sometimes in the form of a current summed over a time period. In this implementation, jitter in the timing pulses directly affects the performance, see Fig. 10.55 and Sect. 7.5.

Asymmetry in the current feedback path, see Fig. 10.56, stems from differences in rising and falling edges. In this example the data pulse of two successive symbols "1" contains less charge than two single symbols. This problem of a non-return-to-zero (NRZ) pulse sequence can be circumvented by a "return-to-zero" implementation technique. Now every pulse, with or without successor, will return to a zero-value. In a differential design a similar technique is possible. More advanced schemes employing dynamic element matching have been published, e.g., in [335]. Also a switched-capacitor digital-to-analog converter can be used in a time-continuous environment. The opamp forming the summing node now has to handle a combination of continuous and switching signals.

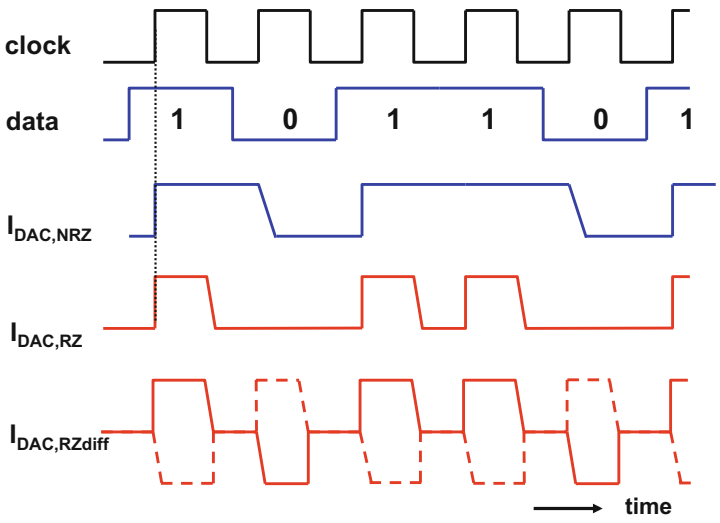


Fig. 10.56 Asymmetry in the digital-to-analog conversion pulses $I_{DAC,NRZ}$ can be mitigated with a return-to-zero technique $I_{DAC,RZ}$ or a differential return to zero technique $I_{DAC,RZdiff}$

10.5.4 Excess Loop Delay in Time-Continuous Sigma-Delta Conversion

A major problem in time-continuous sigma-delta conversion is the delay from the sample pulse to the actual feedback signal. This excess loop delay is caused by comparator delays and time loss in the digital-to-analog converter. In Fig. 10.57 (top) an ideal second order sigma-delta converter with loop delay is shown. Assuming that the comparator does not create any delay $\tau_e = 0$ the path, the ideal transfer function is described with the ideal coefficients $\mathbf{a}_1$ and $\mathbf{a}_2$:

$$\frac{Y(s)}{X(s)} = \frac{\mathbf{a}_1\mathbf{a}_2}{s^2 + \mathbf{a}_2 s + \mathbf{a}_1\mathbf{a}_2}$$

The excess delay is modeled by the function $f_{\tau e}(s) = e^{-s\tau_e}$. This function is approximated with the help of a Taylor series: $e^{-s\tau_e} \approx 1 - s\tau_e + s^2\tau_e^2/2$. Now the transfer function is

$$\frac{Y(s)}{X(s)} = \frac{a_1 a_2}{s^2 + a_2 s f_{\tau e}(s) + a_1 a_2 f_{\tau e}(s)} \approx \frac{a_1 a_2}{s^2(1 - a_2\tau_e + a_1 a_2 \tau_e^2/2) + s(a_2 - a_1 a_2 \tau_e) + a_1 a_2}$$

The terms for $s^3, s^4, \ldots$ are neglected as their coefficients are small. If the coefficient $a_2 = \pi f_s$ (the unity-gain frequency of the second integrator is close to half of the sample rate) and the excess delay is 1/6 of a sample period, then $(1 - a_2\tau_e) \approx 0.5$. It is obvious that the transfer function is not accurately implemented.

Figure 10.57 (bottom) shows the compensation scheme: an additional path weighted with a_e is added to the transfer:

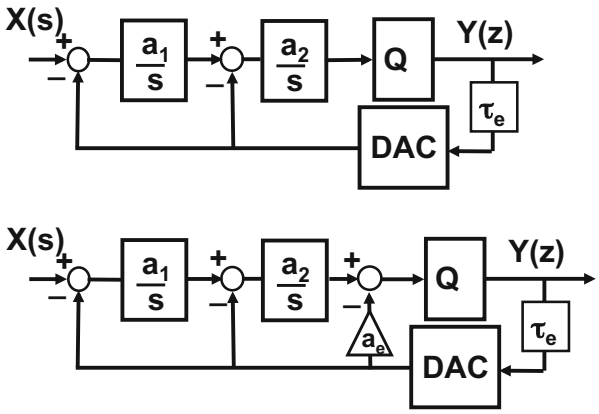


Fig. 10.57 The delay in the quantizer and the digital-to-analog converter τ_e causes excess loop delay. An additional feedback path a_e compensates for this delay

$$\frac{Y(s)}{X(s)} \approx \frac{a_1 a_2}{s^2(1 - a_2\tau_e + a_1 a_2 \tau_e^2/2 + a_e) + s(a_2 - a_1 a_2 \tau_e) + a_1 a_2}$$

This transfer function approximates the ideal second order function if all terms match the terms of the ideal transfer. Equations $1 - a_2\tau_e + a_1 a_2 \tau_e^2/2 + a_e = 1$, $(a_2 - a_1 a_2 \tau_e) = \mathbf{a}_2$ and $a_1 a_2 = \mathbf{a}_1 \mathbf{a}_2$ yield

$$\begin{aligned} a_1 &= \frac{\mathbf{a}_1}{1 + \mathbf{a}_1 \tau_e} \\ a_2 &= \mathbf{a}_2(1 + \mathbf{a}_1 \tau_e) \\ a_e &= \mathbf{a}_2 \tau_e (1 + \mathbf{a}_1 \tau_e/2) \end{aligned}$$

For very fast time-continuous sigma-delta modulation this compensation is absolutely necessary, more analyses are found in [337].

10.5.5 Meta-Stability

In high-speed conversion systems the sample rates are pushed to 10 Gs/s and higher. At these frequencies the time for making a decision and translating in a analog output value is in the order of tenths of picoseconds. These periods are so small that the time constant of an optimally designed comparator or digital latch cannot be ignored. With insufficient decision time, the probability of meta-stability becomes an issue. Figure 10.58 shows a second order loop where on two positions sources are added that inject bit errors due to meta-stability of the comparator and the digital-to-analog converter [338].

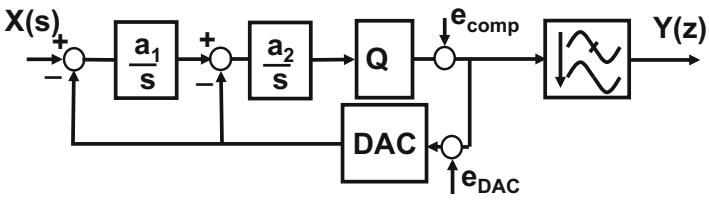


Fig. 10.58 Meta-stability can occur in the comparators and in the digital-to-analog conversion

If bit errors occur that can be modeled by an injection source directly after the quantizer, the errors will be part of the output signal and of the loop feedback signal. The reconstruction of the output data stream and inside the loop is based on the same signal. Consequently the bit error will create some local noise.

The other scenario implies that bit errors occur in the digital-to-analog converter. Now the sigma-delta modulator processes another signal than the reconstruction filter and the meta-stability creates much more severe problems. Therefore specifically attention must be paid to the design and lay-out of the digital-to-analog converter. Bolatkale et al. [339] uses a 15-level quantizer, where every comparator directly switches one current source of the digital-to-analog converter.

10.5.6 Latency

The latency of a sigma-delta converter is often seen as one of its main disadvantages. Yet, this argument has to be judged carefully. The total delay between the applied input signal and the resulting Nyquist-format digital output is composed of the delay from input to output of the modulator, plus the delay in the succeeding down-sample filter. The first component is mainly determined by the delay in the analog loop filter, which is dominated by the first pole. The subsample filter, see Sect. 10.4.9, will convert the oversampled bit-stream from the sigma-delta modulator into digital PCM code at a Nyquist-rate sample rate. The delay in the down-sample filter can run into many clock periods, especially if there is a requirement on linear phase of FIR filters.

10.5.7 Filter Implementations

A continuous-time modulator with a feed-forward structure and a simple 1-bit quantizer and feedback digital-to- analog converter results in rather optimum performance [310]. Even extreme frequency ranges come in reach as shown by Bolatkale et al. [339]. Figure 10.59 gives an impression of opamp topologies needed for 125 MHz bandwidth time-continuous sigma-delta modulators. The chip

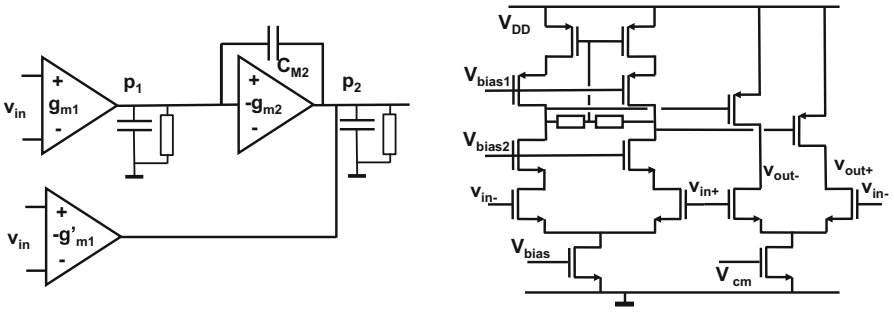


Fig. 10.59 *Left*: topology of the filter opamp, and *right*: transistor schematic. In 45-nm CMOS the gain is 35 dB at a UGBW of 8 GHz and a power consumption of 23 mA [339]

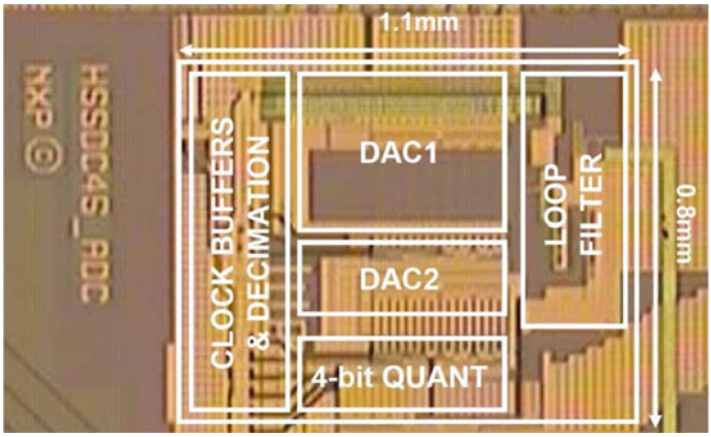


Fig. 10.60 Chip photograph of a 125 MHz bandwidth 65 dB SNDR time-continuous sigma-delta modulator [339], "DAC1" is the overall feedback, while "DAC2" implements the excess loop delay path. Courtesy: M. Bolatkale, NXP

photograph in Fig. 10.60 shows the large area occupied by the digital-to-analog conversion. The filter topologies are often dictated by the need to have a second summation before the quantizer. Filters based on transconductances or OTA's deliver current as an output, which greatly simplifies the design of the summation point. In combination with capacitors, standard filter characteristics can be easily set-up and simulated. Some designs use additional feedback paths to realize zeros. In complex sigma-delta converters the entire toolbox of poly-phase filtering is applied to create the required transfer characteristics and quantization error suppression.

10.6 Time-Discrete and Time-Continuous Sigma-Delta Conversion

Figure 10.61 summarizes a number of architectural choices and consequences for the designer of sigma-delta converters. In most design projects the overall system specifications determine the desired conversion bandwidth. Based on this bandwidth Fig. 10.10 allows to determine the required oversampling factor and the resulting sample rate in relation to the filter order.

In first instance it seems that the difference between time-discrete and time-continuous sigma-delta conversion is limited to the implementation of the filter. Yet this choice has some major consequences, see Table 10.3. From an accuracy point of view the optimum choice for realizing time-discrete filters in standard CMOS is the switched-capacitor technique. The poles and zeros are in the frequency domain determined by the capacitor ratio and the sampling frequency. An accuracy of 0.1 % is here achievable. Switched-capacitor sigma-delta converters are therefore a natural choice for realizing cascaded structures as the matching of filter characteristics between the loop filter and the digital filter in the noise compensating branch is essential. In time-continuous filters, the filter sections can be designed with mutually matched g_m/C sections which will guarantee a well-defined shape of the curve over process variations. The absolute position, however, is determined by the actual process parameter values and can vary $\pm 10\,\%$. Breems et al. [324] shows that a calibration mechanism can make a time-continuous filter match over a certain frequency range to the characteristics of the digital noise filter, although a time-discrete digital filter can only match a time-continuous filter over a limited part of the frequency band.

The choice for a time-continuous loop filter eliminates the sampling mechanism at the input of the sigma-delta modulator and shifts the sampling to the quantizer circuit. This configuration allows to use the time-continuous filter as an alias filter. This feature is efficient from a system point of view especially with higher-order filters.

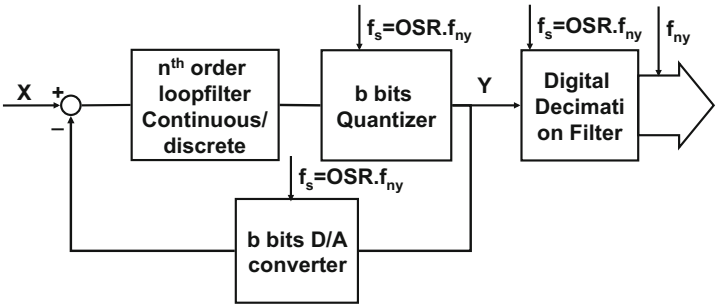


Fig. 10.61 Several design choices play a role during the implementation of a sigma-delta converter: the oversampling factor OSR, the levels in the quantizer, the type, and order of filter

Table 10.3 Comparison of time-discrete and time-continuous sigma-delta modulation

	Time-discrete	Time-continuous
Filter arrangement	Often CIFB to stabilize the second order	Often CIFF to reduce demands on second and higher integrator stages
Sample rate	Filter scales to the sample rate, range of sample rates allowed	Sample rate is fixed within a window to fit to the filter shape
Jitter	Similar to normal sampling	When using current feedback an order more sensitive
Relative positions of poles and zeros in filter	Defined by capacitor ratios	Matching of capacitor ratios and ratios between resistive elements
Absolute value of filter parameters	Capacitor ratios and sample rate	Filter curve modulated by absolute spread of parameters
Relative and absolute accuracy of poles and zeros	<0.1 %, 0.1 %	0.5 %, 15 %
Single loop	Practical implementations are second order	Up to fifth order
Higher-order noise shaping	Cascade of lower order sigma-delta modulators	Increase filter order
Linearity	Summation nodes and integrators	Mainly in first integrator
Anti-alias filter	Extra time-continuous filter required	Partly or completely implemented by loop filter
Out-of-band signals (blockers)	Can be modulated back	Suppressed by loop filter
Timing	Can be chosen in quiet phase	Runs continuously
Opamp constraint	Runs at sample rate	Determined by filter
Power	In fast opamps	In linearity and noise of first integrator

In various communication systems one of the major system requirements is the out-of-band interferer suppression (blocker suppression). Blocker signals are neighboring signals in the communication band with maximum amplitude. In many Wifi and mobile phone systems these are a major worry. After IF-demodulation blockers appear somewhere above the desired band. Their amplitude causes opamps to saturate. In time-discrete converters the sampling of an out-of-band interferer takes place before filtering. The frequency band in which the alias products return depends on the position of the interferer frequency with respect to the sample rate. The time-continuous filter first reduces the amplitude of an interferer signal before the sampling process takes place [340, 341]. A similar argument holds for signals that are injected through on-chip substrate coupling or EMC.

The power of the active elements that implement the time-continuous filter is largely linked to the filter bandwidth and noise/linearity specification. A switched-capacitor time-discrete sigma-delta modulator shifts charges around in its analog

circuits. These charges are transferred from one opamp-capacitor section into another and require sufficient settling for the switched-capacitor circuit. The processing speed of these sections is equal to the sampling rate. In first order the opamp bandwidth may exceed the sampling rate. Multiplexing opamps over a few filter sections is possible, although memory effects become an issue. A significant amount of power is needed in the opamps to allow a fast settling. Many authors therefore argue that a power advantage of time-continuous sigma-delta converters exists for the same filter specification.

Circuit designers want to increase the oversampling ratio for optimum sigma-delta conversion. However, a large bandwidth-to-sample rate ratio requires a filter capacitor ratio of a similar ratio. Due to matching requirements and parasitic components there is a minimum-size for the capacitor value. This implies that the counterpart capacitor in the filter is potentially large and requires a lot of power in its driving opamps. This dilemma is less pronounced in time-continuous converters. Also the on-resistance of the switches become important to consider if large capacitors are needed in the switched-capacitor filter. In time-continuous filters it is mainly the bandwidth and dynamic range of the wanted signal that sets the requirements for the filter design. An exception is the input summing circuit, where the analog input signal meets the time-discrete feedback signal. If both are in the voltage domain, this node will experience large voltage excursions and is consequently sensitive to distortion. The summing circuit and successive stages amplify the remaining signal, so the thermal and $1/f$ noise performance of the input circuit is most critical. The noise and distortion arguments make that the input summing circuit is the most challenging part of the time-continuous design.

An advantage of shifting charges in a switched-capacitor circuit is that the result of this process, the total amount of transferred charge, is independent of the timing. Shorter or longer sample periods due to jitter will modulate on the settling tail only, but not affect the amplitude to a serious degree. In a time-continuous sigma-delta converter often the feedback path is implemented in the current or voltage domain. The feedback signal equals the integral of the voltage or current over the sample period. This means that any jitter or asymmetry in the timing edges of the feedback pulse is heavily impacting the performance. Figure 7.55 shows the impact of jitter on pulse-width modulated signals. This impact of jitter in a time-continuous converter is its major limitation. As the jitter is multiplied by the amplitude of the feedback digital-to-analog converter, the single bit implementations (with the largest feedback amplitude) are most sensitive.

Both approaches, time-discrete and time-continuous, have advantages and disadvantages. Some research approaches combine elements of both techniques in the feedback path [324, 342] or apply an FIR digital-to-analog converter to limit jitter influence [343].

Figure 10.62 compares the time-continuous and switched-capacitor sigma-delta converters that have been published on the ISSCC. Time-discrete switched-capacitor sigma-delta modulators are more present in the high resolution region, while time-continuous converters push forward towards higher bandwidths. In Fig. 10.63 the

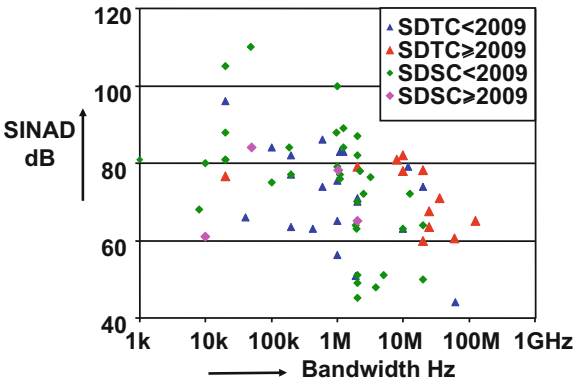


Fig. 10.62 Bandwidth and SINAD comparison for time-continuous and switched-capacitor sigma-delta converters as published on the international solid-state circuits conference (from: B. Murmann, "ADC Performance Survey 1997–2015," [Online]. Available: http://www.stanford.edu/~murmann/adcsurvey.html)

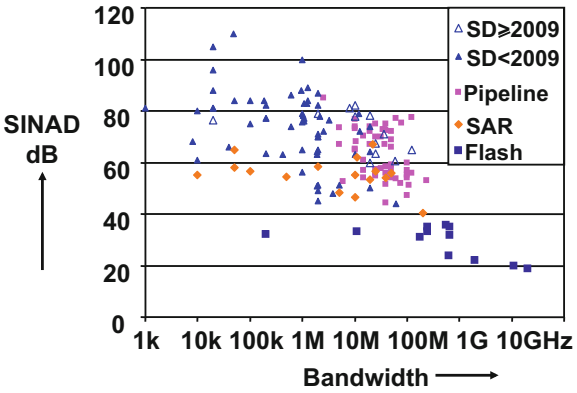


Fig. 10.63 Bandwidth and SINAD comparison for sigma-delta converters, pipeline, flash, and SAR converters as published on the international solid-state circuits conference (from: B. Murmann, "ADC Performance Survey 1997–2015," [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html)

comparison is extended with other converter architectures. It is obvious that in the last years time-continuous sigma-delta conversion enters the domain of pipeline conversion [310, 339, 344].

Example 10.8. Discuss the various noise sources in a sigma-delta converter.

Solution. For many applications quantization distortion/noise can be strongly reduced by a proper choice of oversample ratio and filter order.

Jitter and thermal noise are the main contributors. Jitter is especially dominant in fast one-bit feedback loops. It must be minimized by proper design of the timing path and a high-quality sample-rate source. In many architectures thermal noise is most critical at the first summation point. Linearity demands do not allow excessive gains, and the dynamic range has to be sufficient to accommodate the sum of input and feedback signals.

Very low-bandwidth converters may suffer from a significant $1/f$ contribution. Chopping the bias paths will largely remove this term.

Finally some contributions from meta-stability in the comparator and residual components from digital-to-analog converters with data-weighted averaging can contribute to the total noise budget.

The initial design of a typical SINAD = 80 dB sigma-delta converter running at several tenths of MHz bandwidth will allocate the noise contributions roughly in three parts: 30 % for jitter and the same for thermal noise and the rest for the other components.

10.7 Multi-Bit Sigma-Delta Conversion

In sigma-delta conversion the signal-to-noise improvement by oversampling and subsequent noise shaping is huge, at times over 100 dB. The signal-to-noise gain obtained by a multi-bit quantizer (6 dB/bit) offers from that perspective little advantage over a single-bit quantizer. In addition a single-bit quantizer uses only two levels for the feedback path, which is inherently linear. This last argument enabled designers to improve harmonic distortion performance beyond the levels set by multi-bit quantization with their inherent mismatches, gradients, etc., causing non-linearities. Yet multi-bit quantizers with output wave forms as in Fig. 10.64 enjoy an increasing popularity [324, 345, 346]. The major motivation for multi-bit quantizers comes from the implementation issues that arise with a single-bit quantizer. The discussion can be summarized as follows:

- A single-bit quantizer generates a lot of noise in the loop. This noise has to be treated in a linear way, because non-linearities will cause mixing products in the desired band. The lower noise level of a multi-level converter alleviates this requirement.
- The starting signal-to-noise level of a single-bit quantizer before the noise reduction is low. Every bit of extra quantization lowers the level of the overall noise spectrum and increases the signal-to-noise ratio (*SNR*). Table 10.4 is based on Table 4.2.

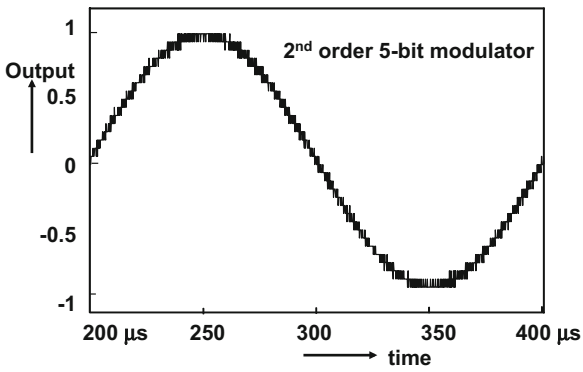


Fig. 10.64 The reconstructed output wave form of a 5-bit quantizer in a sigma-delta modulator

Table 10.4 Simulated noise level reduction for multi-level quantization

Resolution	ΔSNR
1→2	7.0 dB
2→3	6.2 dB
3→4	6.1 dB
4→5, etc.	6.0 dB

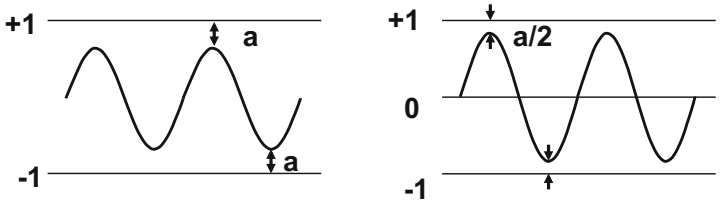


Fig. 10.65 Signals close to overload in a 2-level and 3-level quantizer

- The summation node has to deal with a multitude of requirements. Single-bit quantizers cause large and steep feedback pulses. A multi-bit approach relieves this issue for the summing node and to a lesser extent for the succeeding filter stages.
- The sensitivity to jitter and timing asymmetries is lower for smaller steps of a multi-level digital-to-analog converter in the feedback loop.
- The swing of the residue signal in the first integrator of a multi-level quantizer is reduced, so a larger integrator gain is possible reducing distortion.
- A multi-level quantizer requires more comparators, more digital hardware, and a more extensive digital-to-analog converter, which translates into more power.
- The gain factor in a single-bit quantizer has to be determined by comparing the input and output power. In a multi-level quantizer the gain factor is much better defined. The entire definition of filter and stability is better controlled and less gain margin has to be built into a multi-level loop design.
- The overload level of a multi-level modulator improves [347] and thereby increases the signal-to-noise ratio. In Fig. 10.65 the 2-level quantizer reconstructs a sine wave with pulses switching between $+1$ and -1. For proper operation, the sine wave amplitude is limited to a fraction $\pm(1-a)$ of the range defined by the levels $+1$ to -1, where $a \approx 0.3$. van Veldhoven et al. [347] assumes that in a multi-level quantizer in the most positive and most negative interval the same distance between the signal and a single step of the feedback path is required for proper operation. If this distance is a on either side for two levels, it is $a/2$ for a three levels between ± 1, and $a/3$ for four levels between -1 and $+1$. In general the signal swing as a fraction of the total range for a quantizer with n_q quantization levels is

$$\alpha_{o,nq} = \frac{n_q - 2a}{n_q - 1} \tag{10.39}$$

Figure 10.66 shows a graph of this formula indicating that some 2.5 dB can be gained in signal amplitude.

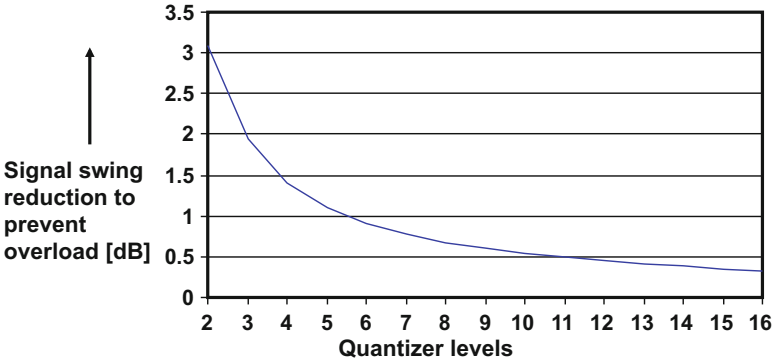


Fig. 10.66 The required signal reduction to prevent overload reduces with increasing number of quantization levels

The multi-bit quantizer is often designed as a flash converter. However, in a flash converter every bit of resolution doubles the power, area, and loading. Many papers have been published over the last years, and almost any Nyquist converter has been tested as multi-bit quantizer: subrange converters, pipeline, cyclic, and very often successive approximation. All these replacements directly cause a problem as their larger latency causes stability issues in the loop. Ranjbar et al. [348] applies a form of excess delay compensation in a successive approximation converter. Others apply self-timing to reduce the latency. Expected problems with voltage headroom have triggered VCO-based quantizers in the sigma-delta loop. The quantizers as discussed in Sect. 8.9.3 show considerable non-linearities, that can jeopardize a good performance.

As a sigma-delta modulator is a feedback system, the quality of the entire conversion is determined by the properties of the feedback path. When the quantizer levels are not accurate there will be a minor impact on the quantization noise suppression. However, the linearity of the digital-to-analog converter in this path limits the overall achievable linearity. When the quantizer is two-bit and consequently the digital-to-analog converter is two bit, this digital-to-analog converter must exhibit a 16-bit (timing) accuracy to be able to achieve a 16-bit resolution sigma-delta conversion. The main challenge of multi-level sigma-delta conversion is therefore the design of a digital-to-analog converter in the loop that must meet the overall specification. With the introduction of accurate digital-to-analog conversion techniques, such as data weighted averaging [171], see Sect. 7.6.3, the application space for multi-bit quantizers has considerably increased. Yet, the error-shaping in the digital-to-analog converter must be carefully designed in order not to dominate the overall noise reduction. Moreover, the delay in the processing must be kept low. Also the tones that can occur in simple DWA algorithms must be suppressed by means of additional randomizing steps.

See Table 10.5 for an overview of arguments. A provocative (yet instructive) case against one-bit sigma-delta is made in [349].

Table 10.5 Trade-off arguments for 1-bit versus multi-bit sigma-delta modulators.

	1-bit	N-bit
Quantization noise	Maximum	Quantization power reduced by N^2
Jitter	Maximum sensitivity	Jitter power reduced by N^2
Pulse asymmetry	Poor	Better
Power consumption	Low	Increased by 2^N comparators and DAC cells
Linearity	Implicitly good	Need additional measures such as DWA in the DAC

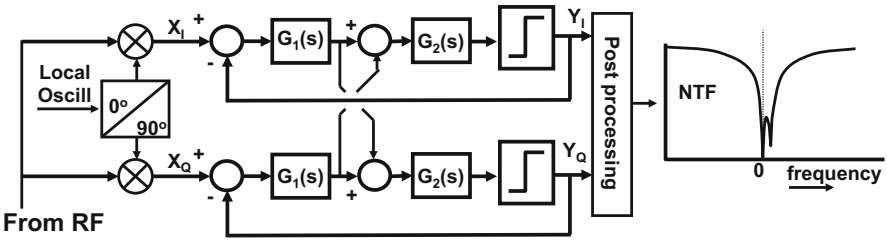


Fig. 10.67 The complex sigma-delta converter consists of two coupled sigma-delta modulators, e.g., [324, 350]

10.8 Various Forms of Sigma-Delta Modulation

10.8.1 Complex Sigma-Delta Modulation

In digital communication systems signals are modulated with an in-phase carrier and its 90° shifted quadrature component. These I/Q signals are converted to the digital domain with a complex sigma-delta converter, Fig. 10.67. Such a converter is built from two identical sigma-delta converters. The input signal from the RF circuits is down-modulated with two phases of the local oscillator. The X_I and X_Q signals are fed in the two sigma-delta modulators. The phase difference between the two signals means that the quadrature signal is a quarter of a carrier period delayed with respect to the in-phase signal. This property allows to form a complex frequency domain where the positive frequency domain is not a mirrored version of the negative side. Moreover, the mutual coupling between the two signal paths allows to define single-sided filter characteristics for the NTF. These properties are beneficial for designing an optimum conversion for various communication systems.

10.8.2 Asynchronous Sigma-Delta Modulation

The asynchronous sigma-delta converter [351] in Fig. 10.68 uses the same topology as the time-continuous converter with the exception of the quantizer. The clocked quantizer is replaced by a non-clocked non-linear element, such as a continuous

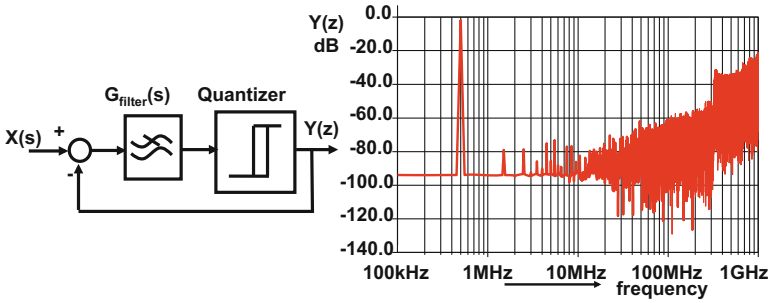


Fig. 10.68 The asynchronous sigma-delta converter and output spectrum

comparator, or a comparator with some hysteresis. There is initially no quantization involved as the comparator delivers a continuous pulse-width modulated signal. So, there is no quantization energy created to be shaped. The loop operates as a kind of multi-vibrator on a self-oscillation frequency. This frequency is determined by the loop filter delay and the hysteresis amplitude. Nevertheless the comparator generates high-order distortion products, that also modulate the oscillation frequency as in an FM-modulator. Some similarity with pulse-density modulation (PDM) is certainly present. Higher up in the spectrum, see Fig. 10.68 (right), these products are visible.

Another way of looking at an asynchronous sigma-delta modulator is as a synchronous modulator running at sample rate, exceeding at least two times the oscillation frequency.

The main problem with an asynchronous sigma-delta modulator is equivalent to more asynchronous systems: application is possible in a small system without clocked interfaces. When the data at some point is processed by a synchronous component, quantization will occur in a similar way as in a level-crossing analog-to-digital converter, Sect. 8.9.1.

10.8.3 Input Feed-Forward Modulator

In the standard topology of a sigma-delta converter the summation node at the input causes many problems. The combination of the full-signal and the feedback signal creates a large resulting amplitude. This signal requires special attention to avoid distortion and saturation effects such as slewing. A solution to this problem is the input feed-forward architecture as shown in Fig. 10.69. The input signal is fed into the chain just before the quantizer. For the 1× feed-forward path this leads to a transfer function ($c = d = 1$):

$$Y(z) = X(z) + \frac{N_Q(z)}{G(z) + 1} \tag{10.40}$$

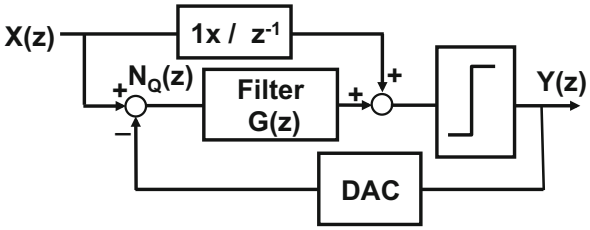


Fig. 10.69 A feed-forward sigma-delta modulator. Vink and van Rens [352] uses a 1× path in the forward branch, while [353] inserts a delay

So the $STF = 1$. The summation node at the input creates the difference between input and output, which ideally contains only the quantization error and no signal. Although this swing can be large as well, the effect on signal distortion is less. Consequently the gain in the first stages can be somewhat larger. This comes at the cost of the introduction of the second summation point.

The topology where the signal is directly fed into the summation node before the quantizer, 1×, has as a disadvantage that in a practical design the signal experiences some time loss by the quantizer and the digital-to-analog converter. This delay creates a phase shift and will lead to larger swings in the filter. In the realization described in [353], a delay is inserted in the forward path to compensate this time loss. This design runs at $V_{DD} = 0.7$ V, with BW = 25 kHz, SINAD = 95 dB, P = 0.87 mW.

10.8.4 Bandpass Sigma-Delta Converter

The dominant direction in the implementation of RF communications systems is the shift of external analog components in the converter. As a result designers are looking for methods to integrate more components of the RF-to-digital interface at the intermediate frequency (IF) level of a super-heterodyne receiver architecture. The antenna signal is first converted to an intermediate frequency to allow sufficient rejection of the image frequency and to filter the wanted signal at a fixed frequency. Depending on the system, the intermediate frequency equals 450–500 kHz for AM radio, 10.7 MHz for FM radio, 33–38 MHz for European television, and tens for MHz for various communication systems.

Sigma-delta conversion allows to directly convert the narrow-band signal at the IF frequency by means of bandpass sigma-delta conversion, Fig. 10.70. Instead of a low-pass filter a bandpass filter will clear the IF frequency band from quantization noise. Engelen et al. [354] reports a sixth order sigma-delta analog-to-digital converter with an FM-radio IF of 10.7 MHz, BW = 200 kHz, ENOB = 10.2, P = 60 mW. A television bandpass converter is shown in [355] and [356] uses a sixth order subsampling bandpass converter for the 2.4 GHz RF band. Harrison et al.

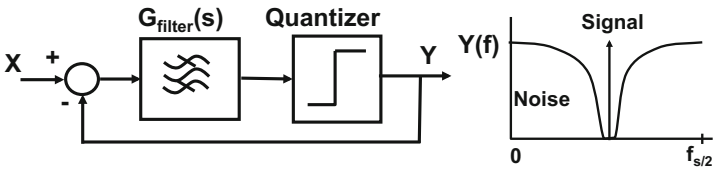


Fig. 10.70 A basic set-up for a bandpass sigma-delta modulator

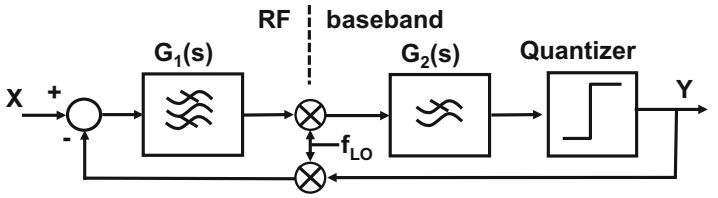


Fig. 10.71 Two mixers divide the architecture in an RF and baseband part in a frequency-translation architecture [360]

[357] implements the bandpass with an LC tank, while [358] uses resonator-type filters to obtain a 24 MHz bandwidth at 200 MHz and 12 mW power. Martens et al. [359] uses an on-chip PLL to generate an 8.8 Gs/s sample rate. The LC-resonator stages create a bandpass at 2.2 GHz. This paper compares some recent bandpass $\Sigma\Delta$ designs. Barkin et al. [161] implements a bandpass digital-to-analog converter at one quarter of the sample rate.

On system level there are doubts about this bandpass concept. The construction of the bandpass filter is tedious. The accuracy of the filter design is in first order related to the pass-bandwidth over center frequency ratio. The quality factor of passive resonators determine the quantization error suppression. Running these filters at high frequencies costs power and area, certainly in a discrete-time realization. In the RF domain dedicated alias filtering has to remove interferers and image frequencies in order to reduce the requirements for the modulator. During the sigma-delta conversion these clean frequency ranges are filled again with quantization power. After the bandpass converter there is another filter needed to remove that unwanted spectrum.

Low-IF or zero-IF topologies with a down-mixer combined in the first stage of the sigma-delta modulator [335, 350, 361–363] provide an alternative. van der Zwan et al. [362] describes a sigma-delta analog-to-digital converter with an IF of 10.7 MHz, BW = 200 kHz, carrier-noise = 79 dB (12.8 ENOB) with a power of 11 mW. Tao and Khoury [360] discusses an intermediate solution between bandpass and a zero-IF down-mixer approach: the frequency translation topology, Fig. 10.71. Two mixers divide the modulator in a high-frequency and a low-frequency part. Smart choices of the filter implementation allow an interesting trade-off, e.g., G_1 provides some bandpass, where the baseband filter G_2 creates suppression.

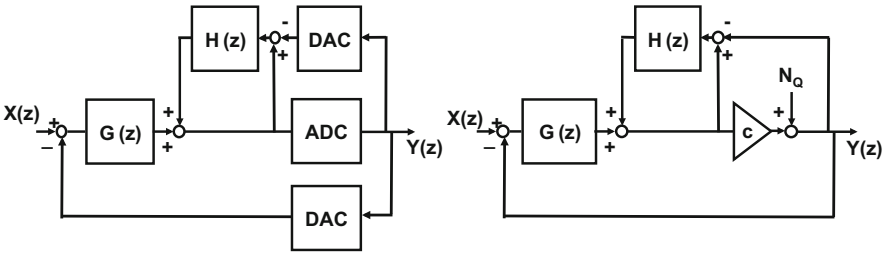


Fig. 10.72 A sigma-delta converter with embedded noise shaper. *Left* is the design and *right* the equivalent scheme [308]

10.8.5 Sigma-Delta Loop with Noise-Shaping

Higher-order noise suppression can be achieved with high-order loop filters or with cascading of sigma-delta modulation loops. Higher-order filters require a higher oversampling ratio to allow the transfer function turning back to a first order behavior around unity gain. The cascaded sigma-delta needs more hardware to implement higher-order noise shaping and depends on the matching between the analog and digital filters.

A third technique combines the advantages of a relative low-order sigma-delta loop and an embedded noise shaper, see Fig. 10.72 (left). The noise shaper is constructed around the low-resolution analog-to-digital converter that serves as a quantizer. The difference between the input of the quantizer and the analog version of the output is fed into a filter $H(z)$ and re-inserted into the loop. This quantizer with noise shaper serves as a modified quantizer in the overall sigma-delta loop. With the help of the equivalent scheme in Fig. 10.72 (right) the transfer is calculated as:

$$Y(z) = \frac{cG}{(1+cG)+H(c-1)}X(z) + \frac{1-H}{(1+cG)+H(c-1)}N(z) \approx X(z) + \frac{1-H}{G}N(z) \tag{10.41}$$

Depending on the time-continuous or time-discrete implementation the signals and operations are in the s or z domain. The approximation is valid for $c = 1$ and $G \gg 1$. The term $(1-H)$ reflects the shaping of the noise in the inner part of the circuit, while the term $1/G$ originates from the overall sigma-delta function. This set-up allows to gain some additional noise suppression without the disadvantages of the other solutions.

10.8.6 Incremental Sigma-Delta Converter

The idea behind sigma-delta conversion is to approximate a time-continuous input signal by many samples of a relatively low-resolution digital-to-analog converter. If the input signal is a wave form where the useful information is only available

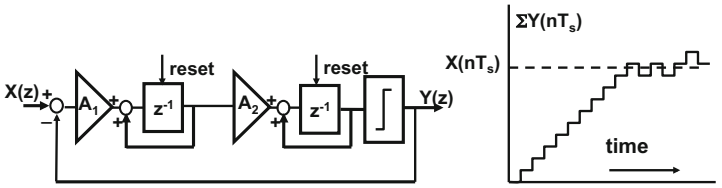


Fig. 10.73 A second order incremental sigma-delta modulator

during a fraction of the sample period, such as the output of a sensor, sigma-delta conversion is not directly applicable. A technique that allows some sigma-delta treatment of periodically available signals is the incremental sigma-delta converter [364], see Fig. 10.73. The main difference with a normal sigma-delta converter is the reset that sets the integrator values back to a starting value. If a signal $x(i) = x_{in}$ for $i = 0 \ldots n$ is applied, the integrators will sum the difference between the input signal and the generated output pulses. After processing n sample periods after a reset, the outputs of the first and second integrator equal:

$$\sum_{i=0}^{n-1}(x_{in} - y(i))$$

$$\sum_{m=0}^{n-1}\sum_{i=0}^{m-1}(x_{in} - y(i)) \tag{10.42}$$

where $y(i)$ is the output for clock cycle i. With sufficient samples processed, the converter will reach an equilibrium. The comparator will be toggling around the output level of the last integrator, generating that pulse density whose average value matches the applied input signal. If the converter reaches a stable pattern, the toggle level of the comparator is assumed to be zero. The resulting transfer function of a converter with one integrator becomes

$$\frac{1}{n}\sum_{i=0}^{n-1} y(i) \approx x_{in} \tag{10.43}$$

The first order incremental sigma-delta converter can be easily read out by summing the positive and negative output pulses. This relatively easy converter circuit demands no stringent requirements on component accuracy or timing and is somewhat faster than linear Nyquist rate converters such as dual-slope. The accuracy improves with a larger number of samples n. The accuracy measured in effective bits improves only slowly with $^2\log(n)$. The main source of concern are sources of time varying behavior, such as drift in the components or interferer signals (unless the total period of operation can be chosen in synchronicity with the interferer).

A second order incremental sigma-delta converter has a transfer function:

$$\frac{2}{n(n-1)} \sum_{m=0}^{n-1} \sum_{i=0}^{m-1} y(i) \approx x_{in} \tag{10.44}$$

In second order incremental sigma-delta converters the two integration steps result in a quadratic relation with n and the accuracy of the total system is proportional to ${}^2\log(n^2)$, which means that for $n = 300 \ldots 600$ a 16-bit resolution is possible. [365] shows a 22-bit third order incremental sigma-delta running at a rate of 15 Hz.

Oike and Gamal [366] discusses the application of incremental sigma-delta conversion in an image sensor.

Exercises

10.1. Compare the heater-thermostat-living room system to a sigma-delta modulator.

10.2. A 6-bit analog-to-digital converter has to convert a signal bandwidth from DC to 50 kHz, but the converter can run up to 200 Ms/s. How much resolution can be obtained? A dither source is available. Can it be used to improve the accuracy?

10.3. A DC-voltage is applied to a first order 1-bit sigma-delta modulator running at 1 MHz. What output frequencies are generated for voltage values of 0,10,... 100 % of the range. What changes if the quantizer is replaced by a 3-bit analog-to-digital converter?

10.4. A first order sigma-delta converter is extended with an integration stage in the feedback path. Draw the STF and the NTF. Is this now a second order sigma-delta modulator?

10.5. An oversample ratio of 64 is available to reach 110 dB SNR gain. How many first order cascade stages are needed to achieve this result.

10.6. The digital-to-analog converter in the feedback path shows a 0.1 % distortion component. Sketch or simulate the effect.

10.7. The second integrator of a second order signal delta modulator shows a 0.1 % distortion component. Sketch or simulate the effect.

10.8. A time-continuous sigma-delta modulator runs at a sample rate of 100 Ms/s. The filter has single poles at 30 and 60 kHz and a zero at 10 MHz. Calculate the NTF and the STF.

10.9. Give an expression for the NTF and STF in a time-discrete sigma-delta modulator running at 100 Ms/s and with poles at 30 and 60 kHz.

10.10. A 1–1 cascaded sigma-delta modulator converts a bandwidth of 10 kHz with an oversampling factor of 512. There is a 2 % mismatch between the gain of the analog filter in the first loop and the digital inverse filter. What will happen?

10.11. A first order sigma-delta modulator runs at a sampling rate of 10 Ms/s. The feedback signal switches between 0 and 1 V. What is the output frequency if the input equals 0.5, 0.6, 0.9 V?

10.12. A first order sigma-delta modulator is part of a 1–1 cascade sigma-delta modulator. Its time-discrete filter is described by $z^{-1}/(1-z^{-1})$. Give a time-discrete description for the digital compensation filter.

10.13. Is it possible to multiplex two identical first order sigma-delta modulators that run on a sample-rate clock and a half-period delayed sample rate.

10.14. A 3rd order time-continuous sigma-delta converter is used to convert a baseband of 2 MHz with an oversample ratio of 64. At what level and frequency will a 255 MHz interferer signal of equal strength as the input signal, show at the output? What happens if this converter is changed into a time-discrete implementation.

10.15. What is the maximum SNR that can be obtained in a 1 MHz bandwidth that is 128× oversampled if the feedback DAC clock jitters 2 ps_{rms}?

10.16. A first order noise shaper is cascaded with a second first order noise shaper. Repeat the derivation in Eq. 10.33 for this situation and define the digital filter.

10.17. A third order time-continuous sigma-delta modulator runs at a sampling rate of 5 Ms/s. A signal band of 20 kHz is required. The comparator and DAC can be assumed ideal, with only the gain uncertainty and the sampling uncertainty of the comparator needs to be taken into account. Set-up a third order filter using capacitors, resistors, and ideal opamps, that gives a maximum SNR ratio in this band.

Chapter 11
Characterization and Specification

Every system designer of analog-to-digital conversion techniques will use specific parameters to determine the usefulness of a converter for the application. The definition of parameters is mostly well described, however, the actual measurement of the parameter can still leave margins for variation [49, 367, 368]. A supplier may restrict himself to the typical value of a parameter, however, also the maximum and minimum values can be part of a specification. Moreover, it may be useful to examine the temperature and supply voltage ranges in which the specified value for a parameter is guaranteed.

The application will determine which parameters of a converter are most relevant. For high-quality audio equipment the distortion is relevant. In communication equipment intermodulation is important as well as the spurious free dynamic range in a certain frequency span. Next to the standard specification points of analog-to-digital converters that are listed in Table 11.1 also a range of secondary qualifications can be taken into account. Examples are phase behavior, package (height, volume, and pin), tolerance to light (e.g., in an optical sensor), available references (internal or external), input impedance, etc.

The correct application of a converter depends crucially on a good characterization, measurement set-up, and parameter extraction.

In this chapter the emphasis is on characterization of a converter, not on testing. During characterization all relevant parameters of a converter are measured under relevant conditions. Even with advanced equipment a characterization takes a lot of time. Testing is done in the factory during mass manufacturing. Testing is about acceptance or rejection of a part and cannot cover the entire characterization space. Choosing the correct test conditions that will detect the potential errors in a chip is a separate field of expertise, that is not covered in this book.

M. Pelgrom, *Analog-to-Digital Conversion*, DOI 10.1007/978-3-319-44971-5_11

Table 11.1 Main characterization parameters of an analog-to-digital converter

Specification	Symbol	Unit
Nominal amplitude resolution	N	1
Sample frequency	f_s	Hz, sec^{-1}
Bandwidth	BW	Hz
Integral linearity	INL	LSB
Differential linearity	DNL	LSB
Monotonicity		
Missing codes		
Harmonic distortion	THD	dB
Intermodulation distortion	IM2,3	dB
Spurious free dynamic range	SFDR	dB
Signal-noise and distortion ratio	SINAD	dB
Signal-noise ratio	SNR	dB
Effective number of bits	ENOB	1
Dynamic range	DR	dB
Jitter	σ_t	ps
Power consumption	P	W
Temperature range	T	°C
Power supply	V_{DD}	V

11.1 Test Hardware

A correct evaluation of a converter starts at the beginning of the chip design. Interfaces to and from the measurement equipment must be defined. These analog or digital drivers and buffers should not interfere with the measurement or jeopardize the signal quality. High sampling frequencies require low-jitter buffers and high-frequency analog output signals require wide bandwidth buffers.

Every experiment starts with a PCB on which the device-under-test (DUT) is mounted. Some more points must be considered when designing a measurement board:

- Analog and digital power supplies and signal sources should be kept separate and only connected together on one single node. Be aware of coupling of earth loops via the mains plug.
- Provide sufficient decoupling: Microfarad electrolytic capacitors for the low-frequencies and metal or ceramic capacitors for the high frequencies. Mount them as close to the package as possible. Of course all available space inside an IC is also filled with decoupling capacitances.
- In the evaluation phase it may be tempting to use a tool that allows to exchange the samples easily. However these tools add a lot of distance between the die and the PCB and therefore add many nano-Henries of inductance. In Fig. 11.2 a technique is shown where the surface mounted device is pushed onto the connection electrodes of the printed-circuit card. This set-up keeps the distances short.

- High-frequency connections must be laid-out keeping in mind that every wire is a transmission line. PCB lay-out packages have options to design wires and surrounding grounding in such a way that a defined impedance is achieved.
- Every signal must be properly terminated close to the measurement device. This certainly holds for digital signals. Non-terminated digital signals will ring and inject spurious charges into the substrate.

A professional measurement set-up for characterizing an analog–digital converter uses a computer to control the set-up and to analyze the measurement results, see also the relevant IEEE standardization documents [48–50]. Many professional evaluation set-ups are constructed with racks of measurement equipment connected by some interface bus. A computer equipped with measurement software will control the equipment, set-up voltages and currents, step through the signal range, and capture the data in a data-logger of several gigabytes storage.

The signal source and the generator for the sample rate have to comply with more stringent specifications than the device under measurement. Modern signal sources are equipped with an extensive user interface, which goes sometimes at the cost of signal distortion and purity. Old "analog" generators are often to be preferred over the modern equipment of the same price level. A well-known method to obtain a high-quality measurement signal uses passive filters, such as the anti-alias filter in Fig. 11.1. This set-up avoids that remaining distortion components, noise of various origins as well as cross-talk of the generators internal processing, disturb the measurement.

The analog-to-digital converter is normally mounted on a load-board: a printed-circuit board adapted for connection to the main measurement equipment, see Fig. 11.2. As a direct coupling of the converter to the measurement equipment may result in long wires, loading, and ground loops, the (digital) side of the converter is buffered near the device with a register file. This buffer will act as a decoupling of signal ringing over the long connection lines. The buffer shields the converter

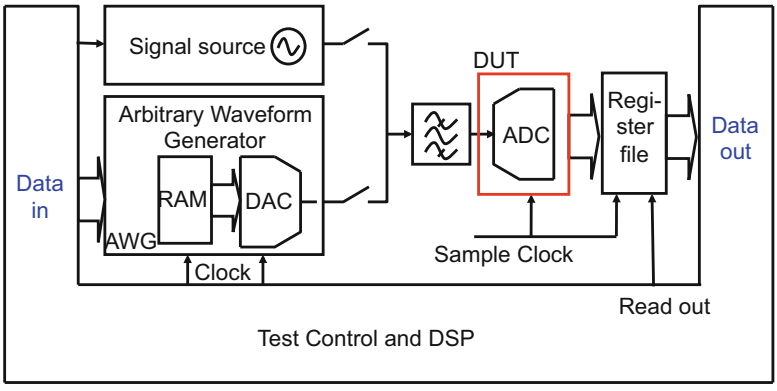


Fig. 11.1 Measurement set-up for digital-to-analog conversion

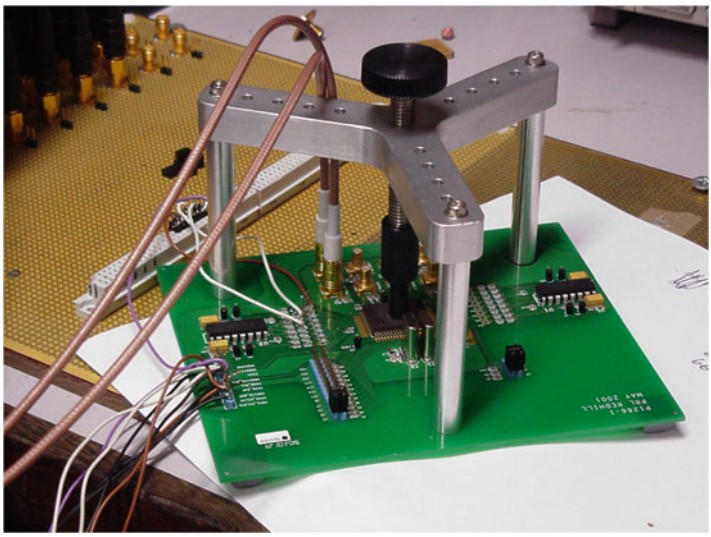


Fig. 11.2 Hardware for measuring an analog-to-digital converter. Courtesy: R. v. Veldhoven

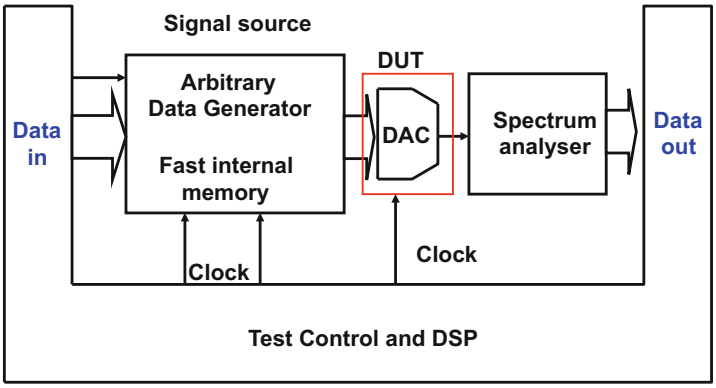


Fig. 11.3 Measurement set-up for a digital-to-analog converter

from the high energies that are associated with driving the measurement bench. In extreme cases the connection between the equipment and the device is made via an optical fiber, so that a perfect electrical separation between the converter and the hardware is achieved. A data storage device (data-logger or data-grabber) will store the high-speed data that comes from the converter. The computer can then in a second phase analyze the data at a convenient speed. The post-processing results in an output as shown in Fig. 11.5.

An important part of the requirements for analog-to-digital measurements apply equally for digital-to-analog converters. Figure 11.3 shows a potential set-up for the measurement. In accordance with the principle of coherent measuring of the

next section, the computer generates and stores a number of data samples in the storage. A cyclic process reads the data at the desired sample rate and feeds the digital-to-analog converter. The required measurement equipment must exceed the specifications of the to-be-tested device. By applying a passive filter, the main component of the output signal can be suppressed so that the measurement equipment only needs to have sufficient resolution for the remaining components. The signal analysis will involve a spectrum analyzer or another form of analog-to-digital conversion.

11.2 Measurement Methods

11.2.1 INL and DNL

A simple way to evaluate the behavior of a converter is to apply a saw-tooth signal to the input. For converters with specifications on absolute accuracy a programmable voltage source is a good choice. Less demanding applications can start with a generator or a self-built circuit. If the saw tooth is sufficiently slow, there will be enough sample moments to determine the DC-parameters as INL, DNL, and monotonicity. If these parameters need to be established at a 0.1 LSB accuracy level, a data storage of 10×2^N samples is necessary.

A saw-tooth signal is not the most critical signal for fast converters and is not easy to generate at high precision. Nyquist analog-to-digital converters with maximum input frequencies ranging from tens of Megahertz into the Gigahertz range require the use of statistical methods. The input signal for measuring dynamic specifications is then preferably a sine wave. This measurement signal can be obtained with a relatively high-quality through the use of passive filters. These fast and relatively accurate methods for determining INL and DNL use the statistical properties of sine waves.

When a full-amplitude sine wave with period T_{sw} is applied to a converter, there is a probability for every code to be hit a number of times. A sine wave will hit more levels in the upper and lower range than in the middle. If the conversion range is defined mathematically between 0 and 1, a full-amplitude sine wave takes the form:

$$y(t) = 0.5 - 0.5\cos(2\pi t/T_{sw}) \tag{11.1}$$

The signal will go from the lowest level to the highest level in half of a cosine period. Δy is a fraction of the range (e.g., 1 LSB) at conversion level y and is called a "data bin." Δt_y is the corresponding fraction of time of the half cosine wave. Δt_y corresponds to the hits in bin Δy while half of the cosine period $(T_{sw}/2)$ corresponds to the total amount of samples. The ratio $\Delta t_y/(T_{sw}/2)$ is now the fraction of hits that end up in bin Δy.

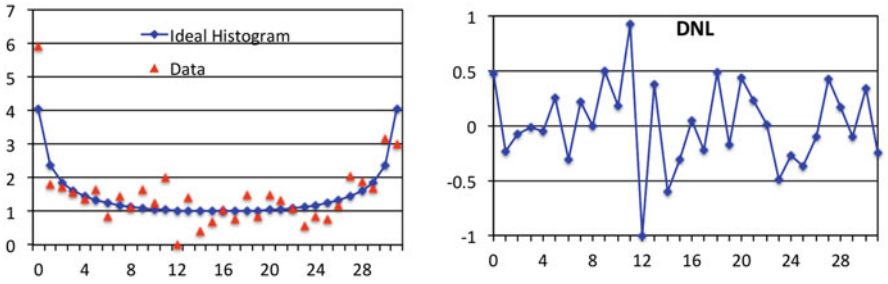


Fig. 11.4 The ideal distribution of hits when a full-scale sine wave is applied to a 5-bit analog-to-digital converter (normalized to 1 for mid-code), measured values, and the corresponding DNL plot (*right*)

$$
\begin{aligned}
t &= \frac{1}{\omega_{sw}} \arccos(1-2y) \\
\frac{dt}{dy} &= \frac{1}{\omega_{sw}\sqrt{y-y^2}} \\
\frac{\Delta t_y}{T_{sw}/2} &= \frac{2}{T_{sw}}\frac{dt}{dy}\Delta y = \frac{\Delta y}{\pi\sqrt{y-y^2}}
\end{aligned}
\tag{11.2}
$$

Δy is chosen as 1 LSB. Figure 11.4 shows a characteristic distribution of the number of hits per level, or binning of levels. A measurement run generates the actual measured distribution of the converted values of a sine wave. This measured distribution is compared to this theoretical curve and the deviations (scaled to the same level) result in an INL and DNL plot, as is shown in Fig. 11.5. This "histogram" method can be used at any signal and sample rate frequency. It provides also information on the linearity problems at higher signal frequencies. The DNL measurement is not optimum as non-monotonicity in this measurement method is not found. Non-monotonicity just changes the DNL value of the corresponding code, the associated step-back is missed. An additional saw-tooth measurement is required.

The calculation above suggests that the input amplitude of the sine wave must accurately match the analog-to-digital converter range. In advanced test packages, routines exist that will allow also amplitudes that extend over the input range. A reconstruction of the input signal is also possible, see Fig. 11.6.

Figures 11.7 and 11.8 give an example of a MatLab code[1] to determine INL and DNL.

Example 11.1. How many samples must be acquired to specify the accuracy of the INL with 0.1 LSB?

[1] Your author is no MatLab wizard, this example comes without any guarantee.

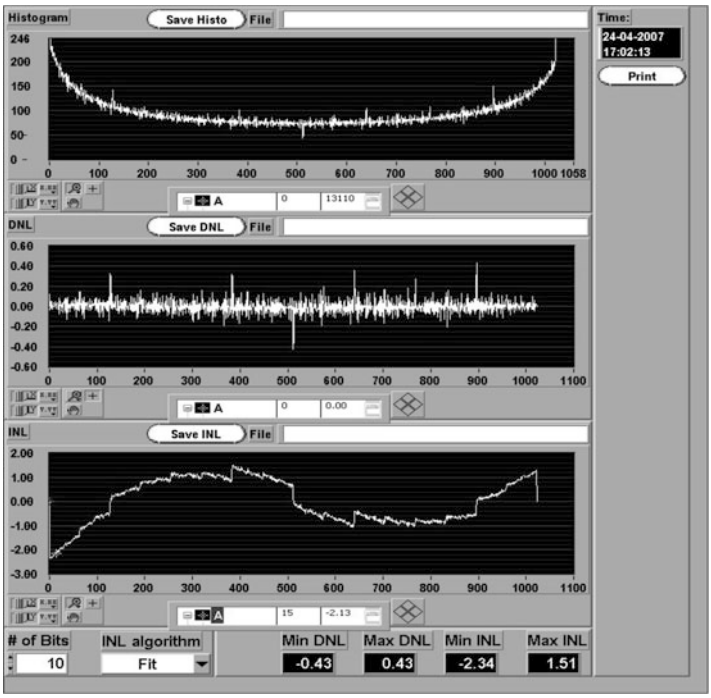


Fig. 11.5 Output of an automated measurement set-up. *Top*: histogram output, *middle*: DNL, *bottom*: INL

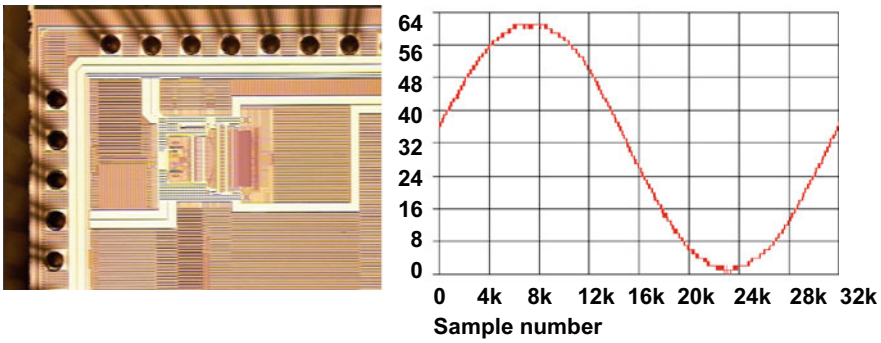


Fig. 11.6 Reconstructed wave form of a 311 MHz signal sampled at 1.44 Gs/s [199]

Solution. In the case of an ideal saw tooth the number of samples in a bin with a size of 1 LSB is determined by the slope of the saw tooth. If the saw tooth rises from minimum reference to maximum reference in N_{ST} sample periods, then the average number of hits per bin will be $N_{ST}/2^N$. In order to obtain an accuracy of 0.1 LSB, N_{ST} must exceed 10×2^N.

```
function [dnl,inl] = dnl_inl_sin(y);
%DNL_INL_SIN
% dnl and inl ADC output
% input y contains the ADC output
% vector obtained from quantizing a
% sinusoid

% Boris Murmann, Aug 2002
% Bernhard Boser, Sept 2002

% histogram boundaries
minbin=min(y);
maxbin=max(y);

% histogram
h = hist(y, minbin:maxbin);

% cumulative histogram
ch = cumsum(h);

% transition levels
T = -cos(pi*ch/sum(h));

% linearized histogram
hlin = T(2:end) - T(1:end-1);

% truncate at least first and last
% bin, more if input did not clip ADC
trunc=2;
hlin_trunc = hlin(1+trunc:end-trunc);

% calculate lsb size and dnl
lsb= sum(hlin_trunc) / (length(hlin_trunc));
dnl= [0 hlin_trunc/lsb-1];
misscodes = length(find(dnl<-0.9));

% calculate inl
inl= cumsum(dnl);
```

Fig. 11.7 On the internet many MatLab procedures for determining the INL and DNL in a data collection can be found. This is a popular page from Berkeley University class EE247

```
% converter model
B = 6;                % bits
range = 2^(B-1) - 1;
% thresholds (ideal converter)
th = -range:range;    % ideal thresholds
th(20) = th(20)+0.7; % error

fs = 1e6;
fx = 494e3 + pi;      % try fs/10!
C  = round(100 * 2^B / (fs / fx));

t = 0:1/fs:C/fx;
x = (range+1) * sin(2*pi*fx.*t);
y = adc(x, th) - 2^(B-1);

hist(y, min(y):max(y));

dnl_inl_sin(y);
```

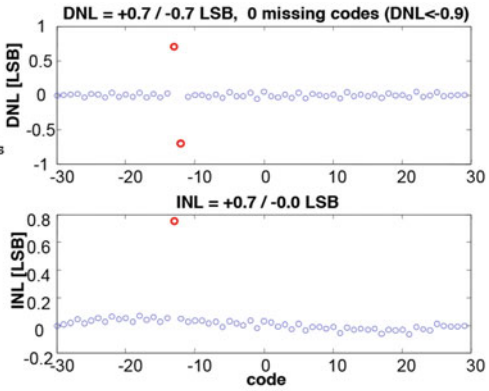


Fig. 11.8 And a test run from the same source

The above calculation for the histogram method allows to determine the minimum number of samples that must be generated to get one hit in the middle bin. There the level corresponds to $y = 0.5$ and

$$\frac{\text{hits in bin at } y = 0.5}{\text{total samples}} = \frac{\Delta t_{y=0.5}}{T_{sw}/2} = \frac{\Delta y}{\pi\sqrt{y - y^2}} = \frac{\Delta y}{\pi/2} = \frac{1}{\pi 2^N/2} \tag{11.3}$$

With a sine wave the number of samples is $\pi/2$ times larger than when applying a saw-tooth signal. To obtain an accuracy of 0.1 LSB in INL and DNL a minimum of $10 \times \pi 2^N/2$ samples is required.

11.2.2 Harmonic Behavior

The same sine waves allow to measure harmonic distortion and related qualities (intermodulation, spurious free dynamic range, etc.) as well as the signal-to-noise ratio.

Fourier transformation of the output sample series allows to generate a frequency diagram, Fig. 11.9. Many Fourier algorithms require that the period in which the data is measured contains both an integer number of signal periods and an integer number of sample periods, see Fig. 11.10. If this condition is not met, the resulting signal will show side lobes, making the interpretation of the Fourier result tedious. This phenomenon is called frequency leakage and is illustrated in Fig. 11.11.

A fast Fourier transform (FFT) is a method to compute efficiently a discrete Fourier transform. This method quantizes the signal components on a grid with frequency bins of size $f_{bin} = 1/T_{meas}$. A time-discrete repetitive signal can be analyzed with a DFT or FFT. If a time-continuous signal is analyzed by means of an FFT algorithm a form of frequency quantizing or discretization takes place, as the frequency components are placed in discrete bins, which can cause errors.

A second pitfall can occur if the sample rate is a simple multiple of the signal frequency. Under stable signal conditions only a limited number of the levels in the conversion process will be used. The evaluation of the converter is based on the repetition of the same limited sequence of measurements, and adds no information on the levels that are missed, see Fig. 11.12.

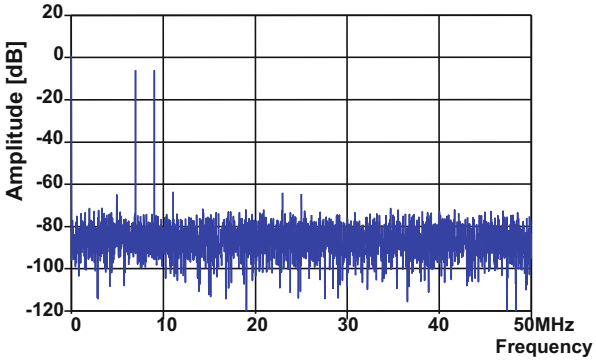


Fig. 11.9 Dynamic measurement of an analog-to-digital converter on intermodulation at $f_s = 100$ Ms/s

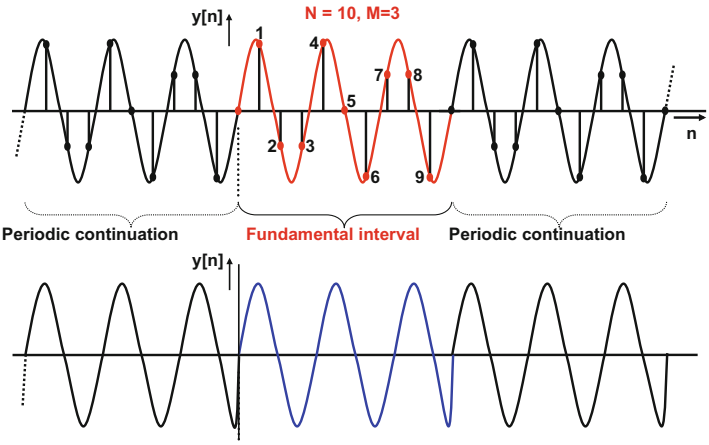


Fig. 11.10 The measured period of a signal is expanded on both sides to enable a Fourier transformation. If the signal and the sample frequency do not fit to the window (*below*) frequency leakage will occur

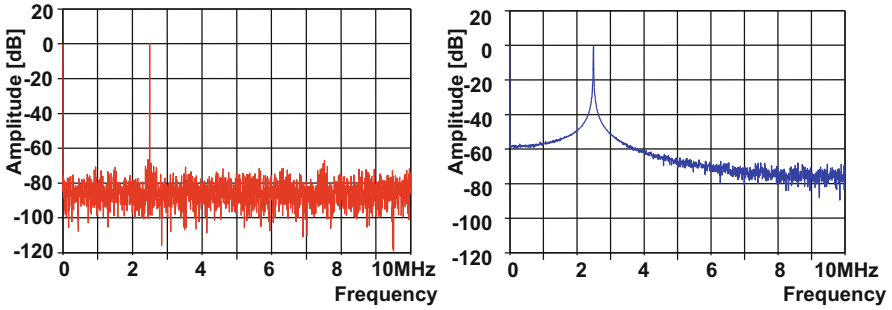


Fig. 11.11 Frequency leakage because of 1 missing sample, *left*: 4000 samples, *right plot*: 3999 samples

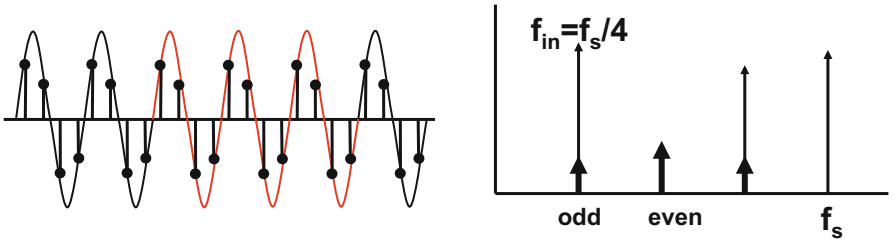


Fig. 11.12 If the sample rate is an integer multiple of the sample rate, a part of the measured samples are simple duplicates of the earlier sequence. In the frequency domain this may lead to the masking of harmonics behind other harmonics or behind the fundamental frequency

Table 11.2 Combinations of sample rates and signal frequencies

f_s Ms/s	f_{signal} MHz	T_{meas} µs	N samples
9	2	1	9
9	2.1	10	90
9.1	2	10	91
9	3.14	50	450
9.01	3.14	100	901

The basic requirement for a good measurement that avoids both problems is called the "coherent testing" condition:

$$T_{meas} = \frac{M_s}{f_s} = \frac{M_{signal}}{f_{signal}} \tag{11.4}$$

Where M_{signal} equals the number of input signal periods and M_s the number of sample periods. If both integers are mutually prime no repetition of measurement sequences will occur. Mutual prime or co-prime means that the largest common divisor of M_{signal} and M_s is 1. The total measurement period is given by T_{meas}.

Another interpretation comes from looking at the smallest common divisor frequency $f_{meas} = 1/T_{meas}$ between the sample rate and the signal. Consequently $f_{signal} = M_{signal} f_{meas}$ and $f_s = M_s f_{meas}$. This frequency is an indication for the achievable accuracy of the measurement.

Example 11.2. An ADC is tested at a sampling speed of 9 Ms/s and a 2 MHz signal frequency. How many unique samples are taken?

Solution. With $f_s = 9$ Ms/s and $f_{signal} = 2$ MHz the smallest common divisor is 1 MHz and $T_{meas} = 1$ µs, so N=9 unique samples. Other options are shown in Table 11.2.

The measurement period is inversely proportional to the frequency resolution or the frequency "binning" of the Fourier transform. It is therefore necessary to choose M_{signal}, M_s, and T_{meas} sufficiently large.

The discrete Fourier transform creates $M_s/2 + 1$ frequency bins of a size $1/T_{meas} = f_s/M_s$. Both bins at 0 and at $f_s/2$ are counted. A spectrum analyzer often provides the option to define the bin size by means of the "resolution bandwidth" parameter. If this value is set, automatically the measurement period will be adjusted.

The energy in the time-discrete signal is distributed over these frequency bins. If energies from different phenomena (e.g., a harmonic component and a folded component) end up in the same bin, the signal strength of these components will add up or extinguish. A finer frequency grid can be obtained by increasing the number of samples by increasing the measurement period T_{meas}.

Figure 11.13 compares spectra taken with 200 and 2000 samples.

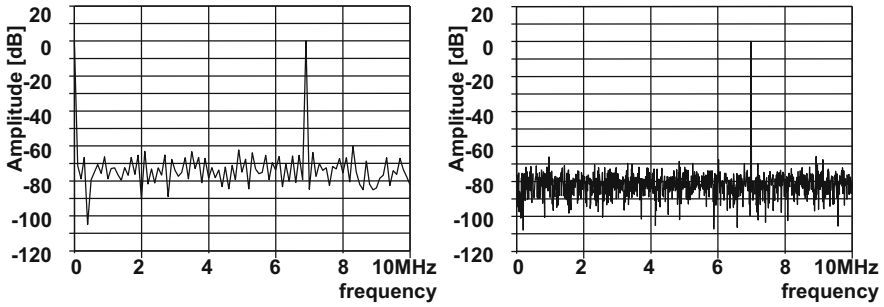


Fig. 11.13 Increasing the measurement period and the number of samples by a factor of 10 reduces the bin size with that factor and lowers the noise floor by 10 dB

Example 11.3. Explain the noise floor in Fig. 11.13.

Solution. The 8-bit analog-to-digital converter has a theoretical maximum signal-to-noise ratio of $1.76 + 8 \times 6\,\text{dB} = 49.8\,\text{dB}$. A measurement and Fourier transform with 200 samples will result in 101 bins. The quantization power in Fig. 11.13 is distributed over these 101 bins, so the power per bin is at a 100× lower power level: at $49.8 + {}^{10}\log(100)\,\text{dB} \approx 70\,\text{dB}$ below the fundamental frequency. The "noise floor" in the spectrum is therefore drawn at approximately $-70\,\text{dB}$. A tenfold increase in samples (and in measurement time) will lead to a ten times lower amount of power per bin. In a spectral plot the noise floor will drop by 10 dB.

Example 11.4. An ADC is measured during 1 ms at a sampling speed of 20 Ms/s, the performance at 3 MHz signal frequency is required. Calculate an appropriate set of measurement conditions.

Solution. With $f_s = 20\,\text{Ms/s}$ and $T_{meas} = 1\,\text{ms}$, a total of $N = 20{,}000$ samples is generated. For a 10-bit ADC this would allow an accuracy of approximately 0.1 LSB. A 3 MHz input sine wave would show 3000 periods, and no coherent conditions can be observed. Changing the input frequency to 2.999 MHz will do. At a measurement period of 1 ms, the spectral resolution (frequency bin) is $1/T_{meas} = 1\,\text{kHz}$.

11.3 Special Measurements

11.3.1 Noise

In a histogram measurement set-up it is relatively easy to perform some special measurements. The noise level (thermal noise, 1/f noise, etc.) can be measured by feeding the input of the converter with a DC value. Of course the noise impedance at the input must be correct. The result as in Fig. 11.14 shows a noise distribution

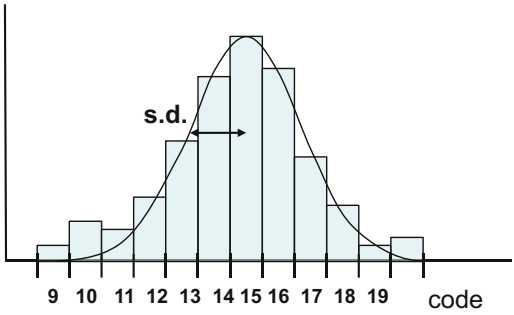


Fig. 11.14 A noise test in a histogram analysis

Fig. 11.15 The bit error rate is measured applying a very slow changing signal to the converter

from which the noise rms amplitude can be found. It is useful to move the DC-value through the input range to check for irregularities.

11.3.2 Bit Error Rate

A second special test measures the bit error rate, Fig. 11.15. In a typical set-up the sample rate is at its maximum and a very slow sine wave is applied to the input. The sine wave changes no more than 1 LSB per step. A data-logger or some simple home-built digital hardware checks whether the output shows code transitions in excess of 1 LSB. These indicate bit errors. The main problem is to guarantee a sufficiently long period of uninterrupted measurement time (air-conditioning switching on/off).

11.3.3 Multi-Tone Power Ratio

In modern communication often multi-tone systems are used to be able to correct path-fading effects. These multi-tone protocols, like orthogonal frequency division multiplexing (OFDM), require special test methods for the conversion process. Figure 11.16 shows the characteristic spectrum for the multi-tone power ratio[2]

[2]Also called: "missing-tone power ratio."

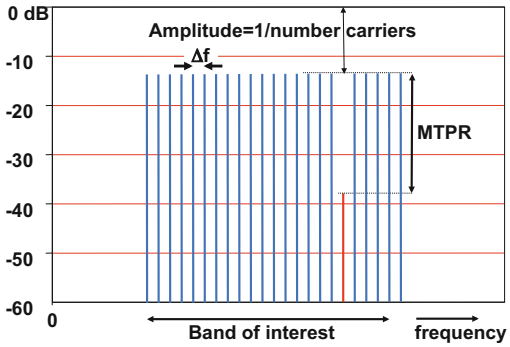


Fig. 11.16 The MTP ratio is measured by generating a pattern of equally spaced tones with one tone missing

measurement and the definition of the MTPR. A pattern of equally spaced tones is generated with one tone missing. Any distortion in, e.g., the digital-to-analog converter in front of the transmitter will create mixing products. As all tones are spaced at Δf their distortion products will appear also on a Δf grid. Some of these intermodulation tones will have a frequency equal to the missing tone and therefore a power component will appear. In Fig. 11.16 the test tone amplitude is limited to 1/number of tones. For a low number of tones there is a reasonable probability that all tones are near their maximum at one moment in time. For larger number of tones this probability becomes so low that a stochastic approximation of the back-off factor is used.

11.3.4 Differential Gain and Differential Phase

Several television broadcast protocols use phase modulation for transmitting, e.g., the color information. For these systems the phase integrity is of course crucial. Differential gain (DG) and differential phase (DP) specify this part of the performance. A staircase signal on which a sine wave modulation is superimposed is applied to the conversion chain. Figure 11.17 shows the ideal result and a trace where differential gain errors and differential phase errors are shown. When the test is executed with a test signal with a constant amplitude sine wave, the differential gain and phase are

$$\mathrm{DG} = \frac{\text{maximum amplitude} - \text{minimum amplitude}}{\text{minimum amplitude}}$$

$$\mathrm{DG} = \text{maximum phase} - \text{minimum phase}$$

Example 11.5. How can a high-speed track-and-hold circuit be measured without having to accurately measure high-frequency output signals?

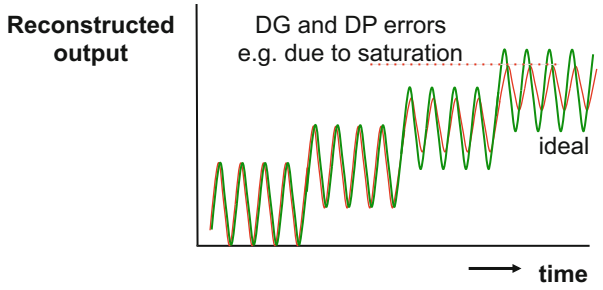


Fig. 11.17 A staircase signal with RF modulation is used to measure differential gain (DG) and differential phase (DP)

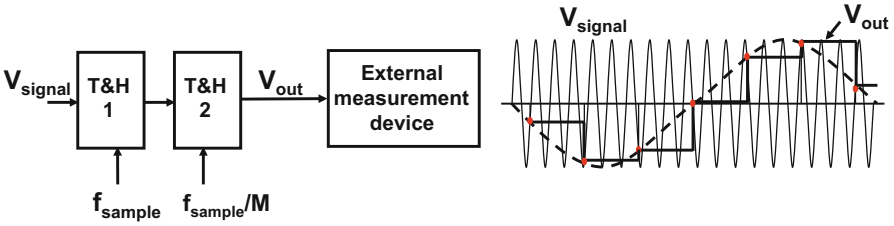


Fig. 11.18 The first track-and-hold circuit is ran at high input and high sample rate. The second device runs a sample rate that is an integer factor slower. The resulting output signal contains signals that correspond to the first track-and-hold output signal and all harmonics

Solution. Subsampling can be used to measure a sampling device such as a track-and-hold circuit. Two devices are cascaded as in Fig. 11.18. The first device is operated at a high sample and signal rate. The second track-and-hold circuit samples the output of the first device at an integer fraction of the sample rate. The original output signal of the first track-and-hold circuit and its harmonics are subsampled to a low output frequency that can be easily measured. The "subsample" method requires quite some skills to interpret correctly the resulting frequency components.

Example 11.6. A very fast sampled data system running at 10 Gs/s must be tested at its extreme frequencies on harmonic distortion, however, there is only a medium-frequency spectrum analyzer available. Design a test set-up using subsampling.

Solution. For testing this sampling system at its extremes, a test signal of 4.987 GHz can be chosen. See Sect. 11.2.2 for an explanation of this incoherent-frequency choice. The second and third harmonics will appear at 9.974 and 14.961 GHz. Due to the sampling mechanism these frequencies will also show up around integer multiples of the sample rate. The second harmonic will be modulated by the 10 Gs/s sample rate as well as the third. Yielding image frequencies at 26 MHz, 4.961 GHz.

This is a standard problem for subsampling. With 16 times subsampling $f_{s,d} =$ 625 Ms/s (=10 Gs/s divided by 16), the three spectral components are brought

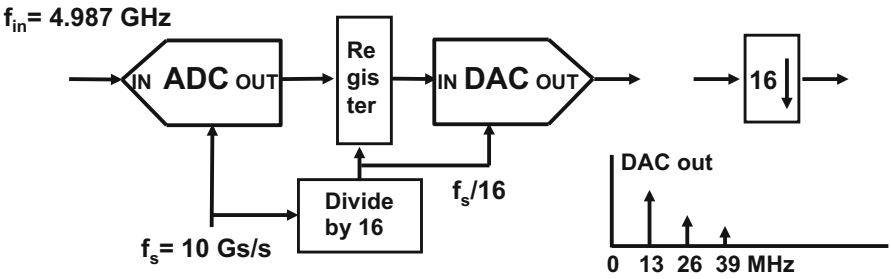


Fig. 11.19 A very fast analog-to-digital converter is tested by subsampling the output data stream. *Right*: symbol for subsampling

into the new Nyquist bandwidth that is half of 625 Ms/s. The signal at 4.987 GHz is modulated down by eight times the new subsample rate (5 Gs/s), yielding a frequency of 13 MHz. In the same way, the second harmonic of the signal appears at 26 MHz and the third at 39 MHz. Figure 11.19 shows a test set-up. Note that the specification for the digital-to-analog converter is very relaxed compared to the analog-to-digital converter: high quality is required in a bandwidth of 40 MHz.

The noise power in the original 5 GHz bandwidth is also modulated into the new $f_{s,d}/2 =$ bandwidth of 312.5 MHz. The total noise power in the Nyquist bandwidths does not change, but the noise power density (power per Hertz) will increase by a factor 16, as the new Nyquist bandwidth is 1/16-th of the old. The noise amplitude will increase by a factor 4.

11.3.5 Self-Testing

In complex systems sometimes forms of self-testing/measuring are necessary. Think of sensor systems that need calibration in places that are difficult to reach. In another example there is a liability aspect to the measurement equipment and the usability of the converter must be established in-situ (e.g., in a drilling head at 2 km below the earth's surface).

Considerations for the implementation of self-test are

- Complexity versus functionality: it may be sufficient to establish correct connectivity of the converter. A simple block wave may be sufficient to test.
- Independence: no test may lead to a positive result because one error has the same effect on the test circuit as on the converter. Using the same reference for the converter as for the test circuit will disable proper detection of reference deviations.
- The cost of error detection: are repair facilities present, or can redundancy lead to a solution (e.g., take a two-out-of-three vote).

- A parametric test can only be performed if somehow accuracy of the test signal is provided. So self-tests create the need for having somewhere a more accurate reference.

Self-testing can be implemented in systems where both a receive and a send chain are present. In a 2.4 GHz transceiver, such a "loop-back" facility feeds a fraction of the transmit power into the receiver. Proper test sequences applied to the digital-to-analog converter input in the send chain allow a functional self-test and also a few parameters can be evaluated.

Another example of self-testing comes from systems where it is impossible to approach the converter. For seismic purposes ships drag large seismic arrays of cables with sensor interfaces. These arrays span several hundreds of meters. Before a measurement is taken the quality of the total interface chain is tested by means of build-in self-test circuits.

It is expected that these professional developments of self-testing in some years will result in a considerable improvement of the performance of self-test methods.

Example 11.7. A 10 bit ADC needs to be tested dynamically at 40 Ms/s in a DSP-based environment.

(a) Determine the minimum measurement time needed to have accessed all codes.
(b) Why is it important that all codes have been accessed? There is 1 ms measurement time available for the FFT.
(c) Determine the input frequency at the Nyquist edge for a good measurement.
(d) What is the number of bins in the FFT?
(e) Determine the approximate noise level seen in the FFT plot.
(f) What can be the technical disadvantage of a long measurement time?

Solution. With 25 ns clock period a ramp signal will take 1024×25 ns=25.6 μs. In case a sinusoidal signal is used 40 μs is needed. Probably the DSP processing will limit this measurement.

All codes need to be accessed to be sure there are no missing codes.

A 1 ms FFT period results in a 1 kHz FFT bin size. For a 40 Ms/s sample rate and a 19.999 MHz input signal, $N = 40.000$ and $M = 19.999$ which numbers are mutually prime.

The number of bins in the FFT is $40.000/2 + 1 = 20.001$ bins of 1 kHz.

The 10-bit converter should have an ideal noise level at $10 \times 6 + 1.76 = 62$ dB. If this noise level is spread over 20.000 bins the level will drop another $10^{10}\log(20.000) = 43$ dB. The total quantization noise level can reach −105 dB, so most likely some thermal noise source will dominate.

A long measurement period allows to perform a detailed FFT, that may reveal more details of the analog-to-digital converters performance.

Exercises

11.1. Compare the advantages and disadvantages of measuring the performance of an analog-to-digital converter by connecting a high-performance digital-to-analog converter to the output or by analyzing the digital output in a signal processor.

11.2. Propose a sine-based equivalent measurement method for a digital-to-analog converter.

11.3. The histogram measurement method uses a sine wave. Set-up a measurement scheme along the same lines using a uniform distributed random signal.

11.4. Can a sigma-delta modulator be measured with a histogram method?

11.5. An 8-bit 20 Ms/s analog-to-digital converter is measured during 0.1 ms with a half-scale sine wave. The result is processed via an FFT. Make a drawing of the expected FFT result.

11.6. A 6-bit ADC needs to be tested dynamically at 4 Gs/s in a DSP-based environment. Determine the minimum measurement time needed to have accessed all codes. There is 40 μs measurement time available for the FFT. Determine the input frequency at the Nyquist edge for a good measurement. What is the number of bins in the FFT? Determine the approximate noise level of the FFT.

11.7. An 8-bit 600 Ms/s analog-to-digital converter is used in a communication system where a spurious free dynamic range of 80 dB in 2 MHz bandwidth is required. What measurement is required.

11.8. An analog-to-digital converter is part of a system-on-silicon. The sample clock is generated on-chip and cannot be accessed separately. Define a method to quantify the jitter of the clock.

Bibliography

1. Seitzer D, Pretzl G, Hamdy NA (1983) Electronic analog-to-digital conversion. Wiley Interscience, New York. ISBN:0-471-90198-9
2. van de Plassche R (1994) Integrated analog-to-digital and digital-to-analog converters. Kluwer Academic Publishers, Dordrecht. ISBN:0-7923-9436-4 (2nd edition. ISBN:1-4020-7500-6, The Netherlands, 2003)
3. Razavi B (1994) Principles of data conversion system design. Wiley/IEEE Press, New York. ISBN:978-0-7803-1093-3
4. Jespers P (2001) Integrated converters D-to-A and A-to-D architectures, analysis and simulation. Oxford Press, Oxford. ISBN:0-19-856446-5
5. Maloberti F (2007) Data converters. Springer, Berlin. ISBN:0-38-732485-2

General Books

6. Carmichael RD, Smith ER (1931) Mathematical tables and formulas. Dover Publications, New York. ISBN:486-60111-0
7. Abramovic M, Stegun IA (eds) (1965) Handbook of mathematical functions. Dover Publications, New York. ISBN:0-486-61272-4
8. Beyer WH (1987) CRC standard mathematical tables, 28th edn. CRC Press, Boca Raton. ISBN:0-8493-0628-0
9. Papoulis A (1965) Probability, random variables, and stochastic processes, Student edn. McGrawHill, New York (4th edition. ISBN:0-073-66011-6, McGrawHill, 2001)
10. Rey WJJ (1983) Introduction to robust and quasi-robust statistical methods. Springer, Berlin. ISBN:0-387-12866-2
11. Sze SM (1981) Physics of semiconductor devices, 2nd edn. Wiley, New York (3rd edition. ISBN:978-0-471-14323-9, 2006)
12. Gildenblat G (ed) (2006) Compact modeling: principles, techniques and applications. Springer, London. ISBN:978-90-481-8613-6
13. ITRS (1994–2013) The international technology roadmap for semiconductors. Last revision was in 2013, now seven focus topics. Updates: http://www.itrs2.net
14. Gray PR, Meyer RG (1993) Analysis and design of analog integrated circuits, 3rd edn. Wiley, New York (4th edition. Wiley, New York. ISBN:0-471-32168-0, 2001)
15. Allen PE, Holberg DR (2002) CMOS analog circuit design, 2nd edn. Oxford University Press, Oxford

2. Sampling

16. Rabiner LR, Gold B (1975) Theory and application of digital signal processing. Prentice-Hall Inc., Englewood Cliffs. ISBN:0-139-141014
17. van den Enden AWM, Verhoeckx NAM (1989) Discrete time signal processing, an introduction. Prentice Hall, Englewood Cliffs. ISBN:0-132-167557

M. Pelgrom, *Analog-to-Digital Conversion*, DOI 10.1007/978-3-319-44971-5

18. Nyquist H (1928) Certain topics in telegraph transmission theory. Trans AIEE 47:617–644; reprinted in Proceedings of the IEEE 90:280–305, 2002
19. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423, 623–656
20. Shannon CE (1949) Communication in the presence of noise. Proc IRE 37:10–21; reprinted in Proceedings of the IEEE 86:447–457, 1998
21. Unser M (2000) Sampling-50 years after Shannon. Proc IEEE 88:569–587
22. Candes E, Romberg J, Tao T (2006) Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inform Theory 52:489–509
23. Hajimiri A, Lee T A general theory of phase noise in electrical oscillators. IEEE J Solid-State Circuits 33:179–194 (1998)
24. Leeson DB (1966) A simple model of feedback oscillator noise spectrum. Proc IEEE 54:329–330
25. Razavi B (1996) A study of phase noise in CMOS oscillators. IEEE J Solid-State Circuits 31:331–343
26. Demir A (2006) Computing timing jitter from phase noise spectra for oscillators and phase-locked loops with white and $1/f$ noise. IEEE Trans Circuits Syst I 53:1869–1884
27. Dalt ND, Harteneck M, Sandner C, Wiesbauer A (2002) On the jit-ter requirements of the sampling clock for analog-to-digital converters. IEEE Trans Circuits Syst I 49:1354–1360
28. Shinagawa M, Akazawa Y, Wakimoto T (1990) Jitter analysis of high-speed sampling systems. IEEE J Solid-State Circuits 25:220–224
29. Zogakis TN, Cioffi JM (1996) The effect of timing jitter on the performance of a discrete multi-tone system. IEEE Trans Commun 44:799–808
30. Clara M, Da Dalt N (2011) Jitter noise of sampled multitone signals. IEEE Trans Circuits Syst II 58:652–656
31. Drakhlis B (2001) Calculate oscillator jitter by using phase-noise analysis. Microwav RF 40:82–90
32. Nathawad LY, Wooley BA, Miller DAB (2003) A 40-GHz-bandwidth, 4-Bit, time-interleaved A/D converter using photoconductive sampling. IEEE J Solid-State Circuits 38:2021–2030
33. McClellan JH, Parks TW, Rabiner LR (1973) A computer program for designing optimum FIR linear phase digital filters. IEEE Trans Audio Electroacust 21:506–526

3. Sample-and-Hold

34. Crols J, Steyaert M (1994) Switched-opamp: an approach to realize full CMOS switched-capacitor circuits at very low power supply voltages. IEEE J Solid-State Circuits 29:936–942
35. Baschirotto A, Castello R (1997) A 1-V 1.8-MHz CMOS switched-opamp SC filter with rail-to-rail output swing. IEEE J Solid-State Circuits 32:1979–1986
36. Dickson J (1976) On-chip high-voltage generation MNOS integrated circuits using an improved voltage multiplier technique. IEEE J Solid-State Circuits 11:374–378
37. Knepper RW (1978) Dynamic depletion mode: an E/D mosfet circuit method for improved performance. IEEE J Solid-State Circuits 13:542–548
38. Abo AM, Gray PR (1999) A 1.5-V, 10-bit, 14.3-MS/s CMOS pipe-line analog-to-digital converter. IEEE J Solid-State Circuits 34:599–606
39. Limotyrakis S, Kulchycki SD, Su DK, Wooley BA (2005) A 150-MS/s 8-b 71-mW CMOS time-interleaved ADC. IEEE J Solid-State Circuits 40:1057–1067
40. Keramat A, Tao Z (2000) A capacitor mismatch and gain insensitive 1.5-bit/stage pipelined A/D converter. In: Proceedings of the 43rd IEEE midwest symposium on circuits and systems, pp 48–51
41. Yang W, Kelly D, Mehr I, Sayuk MT, Singer L (2001) A 3-V 340-mW 14-b 75-MS/s CMOS ADC with 85-dB SFDR at Nyquist input. IEEE J Solid-State Circuits 36:1931–1936
42. Song BA, Tompsett MF, Lakshmikumar KR (1988) A 12-bit 1-MS/s capacitor error-averaging pipelined A/D converter. IEEE J Solid-State Circuits 23:1324–1333
43. Gregoire BR, Moon U (2008) An over-60 dB true rail-to-rail performance using correlated level shifting and an opamp with only 30 dB loop gain. IEEE J Solid-State Circuits 43:2620–2630

44. Wakimoto T, Akazawa Y (1993) Circuits to reduce distortion in the diode-bridge track-and-hold. IEEE J Solid-State Circuits 28:384–387
45. Vorenkamp P, Verdaasdonk JPM (1992) Fully bipolar, 120-MS/s 10-b track-and-hold circuit. IEEE J Solid-State Circuits 27:988–992
46. Sansen W (1999) Distortion in elementary transistor circuits. IEEE Trans Circuits Syst II 46:315–325

4. Quantization

47. Reeves H (1942) A electric signaling system. US Patent 2-272-070, issued February 3, 1942. Also French Patent 852–183 issued 1938, and British Patent 538–860 issued 1939
48. IEEE (1994) IEEE Std 1057–1994 IEEE standard for digitizing waveform recorders
49. IEEE (2000) IEEE 1241–2000, Standard for terminology and test methods for analog-to-digital converters. IEEE Std 1241. ISBN:0-7381-2724-8, revision 2007
50. Tilden SJ, Linnenbrink TE, Green PJ (1999) Overview of IEEE-STD-1241 standard for terminology and test methods for analog-to-digital converters. In: Instrumentation and measurement technology conference, pp 1498–1503
51. Bennett WR (1948) Spectra of quantized signals. Bell Syst Tech J 27:446–472
52. Blachman N (1985) The intermodulation and distortion due to quantization of sinusoids. IEEE Trans Acoust Speech Signal Process 33:1417–1426
53. Oude Alink MS, Kokkeler ABJ, Klumperink EAM, Rovers KC, Smit G, Nauta B (2009) Spurious-free dynamic range of a uniform quantizer. IEEE Trans Circuits Syst II: Express Briefs 56:434–438
54. Lloyd S (1982) Least squares quantization in PCM. IEEE Trans Inform Theory 28:129–137 (transcript from 1957 paper)
55. Max J (1960) Quantizing for minimum distortion. IRE Trans Inform Theory 6:7–12
56. Lipshitz SP, Wannamaker RA, Vanderkooy J (1992) Quantization and dither: a theoretical study. J Audio Eng Soc 40:355–375
57. Wannamaker RA, Lipshitz SP, Vanderkooy J, Wright JN (2000) A theory of nonsubtractive dither. IEEE Trans Signal Process 48:499–516
58. Ceballos JL, Galton I, Temes GC (2005) Stochastic analog-to- digital conversion. In: IEEE midwest symposium circuits systems, pp 855–858
59. Walden RH (1999) Analog-to-digital converter survey and analysis. IEEE J Sel Areas Commun 17:539–550
60. Vittoz E (1995) Low power low-voltage limitations and prospects in analog design. In: van de Plassche RJ (ed) Advances in analog circuit design. Kluwer Academic Publishers, Boston, p 3
61. Dijkstra E, Nys O, Blumenkrantz E (1995) Low power oversampled A/D converters. In: van de Plassche RJ (ed) Advances in analog circuit design. Kluwer Academic Publishers, Boston, p 89
62. Giotta D, Pessl P, Clara M, Klatzer W, Gaggl R (2004) Low-power 14-bit current steering DAC for ADSL applications in 0.13 μm CMOS. In: European solid-state circuits conference, pp 163–166

5. Accuracy

63. Gregor RW (1992) On the relationship between topography and transistor matching in an analog CMOS technology. IEEE Trans Electron Devices 39:275–282
64. Stathis JH, Zafar S (2006) The negative bias temperature instability in MOS devices: a review. Microelectron Reliab 46:270–286
65. Hook TB, Brown J, Cottrell P, Adler E, Hoyniak D, Johnson J, Mann R (2003) Lateral ion implant straggle and mask proximity effect. IEEE Trans Electron Devices 50:1946–1951
66. Drennan P, Kniffin M, Locascio D (2006) Implications of proximity effects for analog design. In: IEEE custom integrated circuits conference, pp 169–176
67. Bianchi RA, Bouche G, Roux-dit-Buisson O (2002) Accurate modeling of trench isolation induced mechanical stress effects on MOSFET electrical performance. In: Technical digest international electron devices meeting, pp 117–120

68. Su K-W et al (2003) A scalable model for STI mechanical stress effect on layout dependence of MOS electrical characteristics. In: Proceedings of the IEEE custom integrated circuits conference, pp 245–248
69. Wils N, Tuinhout HP, Meijer M (2009) Characterization of STI edge effects on CMOS variability. IEEE Trans Semicond Manuf 22:59–65
70. Ge L, Adams V, Loiko K, Tekleab D, Bo X-Z, Foisy M, Kolagunta V, Veeraraghavan S (2007) Modeling and simulation of poly-space effects in uniaxially-strained etch stop layer stressors. In: IEEE international SOI conference, pp 25–26
71. Tuinhout HP, Pelgrom MJM, Penning de Vries R, Vertregt M (1996) Effects of metal coverage on MOSFET matching. In: Technical digest international electron devices meeting, pp 735–739
72. Tuinhout HP, Bretveld A, Peters WCM (2004) Measuring the span of stress asymmetries on high-precision matched devices. In: International conference on microelectronic test structures, pp 117–122
73. Tuinhout HP, Vertregt M (2001) Characterization of systematic MOSFET current factor mismatch caused by metal CMP dummy structures. IEEE Trans Semicond Manuf 14:302–310
74. Pelgrom MJM, Tuinhout HP, Vertregt M (2012) A designers view on mismatch. In: van Roermund AHM, Steyaert M, Baschirotto A (eds) Nyquist AD converters, sensor interfaces, and robustness. Advances on analog circuit design 2012. Springer, New York
75. Lam M-F, Tammineedi A, Geiger R (2001) Current mirror layout strategies for enhancing matching performance. Analog Integr Circ Sig Process 28:9–26
76. Shockley W (1961) Problems related to p-n junctions in silicon. Solid-State Electron 2:35–67
77. Tuinhout HP, Elzinga H, Brugman JT, Postma F (1994) Accurate capacitor matching measurements using floating-gate test structures. In: IEEE international conference on microelectronic test structures, pp 133–137
78. Tuinhout HP, van Rossem F, Wils N (2009) High-precision on-wafer backend capacitor mismatch measurements using a benchtop semiconductor characterization system. In: IEEE international conference on microelectronic test structures, pp 3–8
79. McCreary JL (1981) Matching properties, and voltage and temperature dependence of MOS capacitors. IEEE J Solid-State Circuits 16:608–616
80. Shyu J-B, Temes GC, Yao K (1982) Random errors in MOS capacitors. IEEE J Solid-State Circuits 17:1070–1076
81. Aparicio R (2002) Capacity limits and matching properties of integrated capacitors. IEEE J Solid-State Circuits 27:384–393
82. Tripathi V, Murmann B (2014) Mismatch characterization of small metal fringe capacitors. IEEE Trans Circuits Syst I 61:2236–2242
83. Drennan PG (1999) Diffused resistor mismatch modeling and characterization. In: Bipolar/BiCMOS circuits and technology meeting, pp 27–30
84. Tuinhout HP, Hoogzaad G, Vertregt M, Roovers R, Erdmann C (2002) Design and characterisation of a high precision resistor ladder test structure. In: IEEE international conference on microelectronic test structures, pp 223–228
85. Lakshmikumar KR, Hadaway RA, Copeland MA (1986) Characterization and modeling of mismatch in MOS transistors for precision analog design. IEEE J Solid-State Circuits 21:1057–1066
86. Michael C, Ismail M (1992) Statistical modeling of device mismatch for analog mos integrated circuits. IEEE J Solid-State Circuits 27:154–166
87. Mizuno T, Okamura J, Toriumi A (1994) Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFETs. IEEE Trans Electron Devices 41:2216–2221
88. Forti F, Wright ME (1994) Measurement of MOS current mismatch in the weak inversion region. IEEE J Solid-State Circuits 29:138–142
89. Croon JA, Sansen W, Maes HE (2005) Matching properties of deep sub-micron MOS transistors. Springer, Dordrecht. ISBN:0-387-24314–3

90. Tuinhout HP (2003) Improving BiCMOS technologies using BJT parametric mismatch characterisation. In: Bipolar/BiCMOS circuits and technology meeting, pp 163–170
91. Pelgrom MJM, Duinmaijer ACJ, Welbers APG (1989) Matching properties of MOS transistors. IEEE J Solid-State Circuits 24:1433–1440
92. Brown AR, Roy G, Asenov A (2007) Poly-Si-Gate-related variability in decananometer MOSFETs with conventional architecture. IEEE Trans Electron Devices 54:3056–3063
93. Bastos J, Steyaert M, Roovers R, Kinget P, Sansen W, Graindourze B, Pergoot A, Janssens E (1995) Mismatch characterization of small size MOS transistors. In: IEEE international conference on microelectronic test structures, pp 271–276
94. Stolk PA, Widdershoven FP, Klaassen DBM (1998) Modeling statistical dopant fluctuations in MOS transistors. IEEE Trans Electron Devices 45:1960–1971
95. Andricciola P, Tuinhout HP (2009) The temperature dependence of mismatch in deep-submicrometer bulk MOSFETs. IEEE Electron Device Lett 30:690–692
96. Hook TB, Johnson JB, Cathignol A, Cros A, Ghibaudo G (2011) Comment on (channel length and threshold voltage dependence of a transistor mismatch in a 32-nm HKMG technology). IEEE Trans Electron Devices 58:1255–1256
97. Dadgour HF, Endo K, De VK, Banerjee K (2010) Grain-orientation induced work function variation in nanoscale metal-gate transistors part i: modeling, analysis, and experimental validation. IEEE Trans Electron Devices 57:2504–2514
98. Croon JA, Tuinhout HP, Difrenza R, Knol J, Moonen AJ, Decoutere S, Maes HE, Sansen W (2002) A comparison of extraction techniques for threshold voltage mismatch. In: IEEE international conference on microelectronic test structures, pp 235–240
99. Dahlstrom M et al (2006) Influence of the Ge profile on VBE and current gain mismatch in advanced SiGe BICMOS NPN HBT with 200GHz ft. In: Proceeding of BCTM
100. Tuinhout HP (2005) Electrical characterisation of matched pairs for evaluation of integrated circuit technologies. PhD thesis, Delft University of Technology. http://repository.tudelft.nl/file/82893/025295
101. Takeuchi K, Hane M (2008) Statistical compact model parameter extraction by direct fitting to variations. IEEE Trans Electron Devices 55:1487–1493
102. Cheng B, Roy S, Asenov A (2007) Statistical compact model parameter extraction strategy for intrinsic parameter fluctuation. In: Grasser T, Selberherr S (eds) Simulation on semiconductor processes and devices. Springer, Wien/NewYork, pp 301–304
103. Pelgrom MJM, Tuinhout HP, Vertregt M (1998) Transistor matching in analog CMOS applications. In: International electron devices meeting, pp 915–918
104. Doorn TS, Ter Maten EJW, Croon JA, Di Bucchianico A, Wittich O (2008) Importance sampling Monte Carlo simulations for accurate estimation of SRAM yield. In: 34th European solid-state circuits conference, pp 230–233
105. Vertregt M, Scholtens PCS (2004) Assessment of the merits of CMOS technology scaling for analog circuit design. In: 30th European solid-state circuits conference, pp 57–64
106. Pelgrom MJM (1994) Low-power high-speed A/D conversion. In: 20th European solid-state circuits conference, low-power workshop
107. Kinget P, Steyaert M (1996) Impact of transistor mismatch on the speed accuracy power trade-off. In: Custom integrated circuits conference, pp 333–336

6. Reference Circuits

108. Hilbiber D (1964) A new semiconductor voltage standard. In: IEEE international solid-state circuits conference, digest of technical papers, pp 32–33
109. Widlar RJ (1971) New developments in IC voltage regulators. IEEE J Solid-State Circuits 6:2–7
110. Kuijk KE (1973) A precision reference voltage source. IEEE J Solid-State Circuits 8:222–226
111. Brokaw AP (1974) A simple three-terminal IC bandgap reference. IEEE J Solid-State Circuits 9:388–393
112. Pertijs MAP, Huijsing JH (2006) Precision temperature sensors in CMOS technology. Springer, New York. ISBN:140205257X

113. Song BS, Gray PR (1983) A precision curvature-compensated CMOS bandgap reference. IEEE J Solid-State Circuits 18:634–643
114. Sansen WM, Op't Eynde F, Steyaert M (1988) A CMOS temperature compensated current reference. IEEE J Solid-State Circuits 23:821–823
115. Annema A-J (1999) Low-power bandgap references featuring DTMOSTs. IEEE J Solid-State Circuits 34:949–955
116. Ge G, Zhang C, Hoogzaad G, Makinwa KAA (2011) A single-trim CMOS bandgap reference with a 3σ inaccuracy of $\pm 0.15\%$ from-40 °C to 125 °C. IEEE J Solid-State Circuits 46:2693–2701
117. Banba H, Shiga H, Umezawa A, Miyaba T, Tanzawa T, Atsumi S, Sakui K (1999) A CMOS band-gap reference circuit with sub 1-V operation. IEEE J Solid-State Circuits 34:670–674
118. Petrescu V, Pelgrom MJM, Veendrick HJM, Pavithran P, Wieling J (2006) Monitors for a signal integrity measurement system. In: 32nd European solid-state circuits conference, pp 122–125
119. Petrescu V, Pelgrom MJM, Veendrick HJM, Pavithran P, Wieling J (2006) A signal-integrity self-test concept for debugging nanometer CMOS ICs. In: IEEE international solid-state circuits conference, digest of technical papers, pp 544–545
120. Annema AJ, Goksun G (2012) A 0.0025 mm^2 bandgap voltage reference for 1.1-V supply in standard 0.16-μm CMOS. In: IEEE international solid-state circuits conference, digest of technical papers, pp 364–365
121. Doyle J, Lee YJ, Kim Y-B, Wilsch H, Lombardi F (2004) A CMOS subbandgap reference circuit with 1-V power supply voltage. IEEE J Solid-State Circuits 39:252–255
122. Fayomi CJB, Wirth GI, Achigui HF, Matsuzawa A (2010) Sub 1 V CMOS bandgap reference design techniques: a survey. Analog Integr Circ Sig Process 62:141–157
123. Leung KN, Mok PKT (2002) A Sub-1-V 15-ppm/C CMOS bandgap voltage reference without requiring low threshold voltage device. IEEE J Solid-State Circuits 37:526–530
124. Cabrini A, De Sandre G, Gobbi L, Malcovati P, Pasotti M, Poles M, Rigoni F, Torelli G (2005) A 1 V, 26 μW extended temperature range band-gap reference in 130-nm CMOS technology. In: IEEE European solid-state circuits conference, pp 503–506
125. Gunawan M, Meijer GCM, Fonderie J, Huijsing JH (1993) A curvature-corrected low-voltage bandgap reference. IEEE J Solid-State Circuits 28:667–669
126. Huang HW, Hsieh C-Y, Chen K-H, Kuo B-Y (2008) A 1 V 16.9 ppm/C 250 nA switched-capacitor CMOS voltage reference. In: IEEE international solid-state circuits conference, digest of technical papers, pp 438–439
127. Blauschild RA, Tucci PA, Muller RS, Meyer RG (1978) A new NMOS temperature-stable voltage reference. IEEE J Solid-State Circuits 13:767–774
128. Song H-J, Kim C-K (1993) A temperature-stabilized SOI voltage reference based on threshold voltage difference between enhancement and depletion NMOSFET's. IEEE J Solid-State Circuits 28:671–677
129. Annema AJ, Veldhorst P, Doornbos G, Nauta B (2009) A sub-1V bandgap voltage reference in 32nm finFET technology. In: IEEE international solid-state circuits conference, digest of technical papers, pp 332–333

7. Digital-to-Analog Conversion

130. van de Plassche RJ (1976) Dynamic element matching for high-accuracy monolithic D/A converters. IEEE J Solid-State Circuits 21:795–800
131. Schoeff JA (1979) An inherently monotonic 12 bit DAC. IEEE J Solid-State Circuits 24:904–911
132. Naylor JR (1983) A complete high-speed voltage output 16-bit monolithic DAC. IEEE J Solid-State Circuits 28:729–735
133. Schouwenaars HJ, Dijkmans EC, Kup BMJ, van Tuijl EJM (1986) A monolithic dual 16-bit D/A converter. IEEE J Solid-State Circuits 21:424–429
134. Groeneveld DWJ, Schouwenaars HJ, Termeer HAH, Bastiaansen CAA (1989) A self-calibration technique for monolithic high-resolution D/A converters. IEEE J Solid-State Circuits 24:1517–1522

135. Schouwenaars HJ, Groeneveld DWJ, Termeer HAH (1988) A low-power stereo 16-bit CMOS D/A converter for digital audio. IEEE J Solid-State Circuits 23:1290–1297
136. Jin J, Gao Y, Sanchez-Sinencio E (2014) An energy-efficient time-domain asynchronous 2 b/Step SAR ADC with a hybrid R-2R/C-3C DAC structure.IEEE J Solid-State Circuits 49:1383–1396
137. Meaney RA, Speer RJ (1991) Voltage-switching D/A converter using p- and n-channel MOSFETs. US Patent 5.075.677
138. McLachlan RC, Gillespie A, Coln MlCW, Chisholm D, Lee DT (2013) A 20b clockless DAC with sub-ppm-linearity 7.5 nV/$\sqrt{\text{Hz}}$-noise and 0.05 ppm/°C-stability. In: IEEE international solid-state circuits conference, digest of technical papers, pp 278–279
139. Dingwall AGF, Zazzu V (1985) An 8-MHz CMOS Subranging 8-bit A/D converter. IEEE J Solid-State Circuits 20:1138–1143
140. Abrial A, Bouvier J, Fournier J, Senn P, Viellard M (1988) A 27-MHz digital-to-analog video processor. IEEE J Solid-State Circuits 23:1358–1369
141. Pelgrom MJM (1990) A 10b 50MHz CMOS D/A converter with 75Ω buffer. IEEE J Solid-State Circuits 25:1347–1352
142. Ahuja B (1983) An improved frequency compensation technique for CMOS operational amplifiers. IEEE J Solid-State Circuits 18:629–633
143. Lin C-H, Bult K (1998) A 10-b 250-M sample/s CMOS DAC in 1 mm^2. IEEE international solid-state circuits conference, digest of technical papers, pp 214–215
144. Van Den Bosch A, Borremans M, Steyaert M, Sansen W (2001) A 12 b 500 MS/s current-steering CMOS D/A converter. In: IEEE international solid-state circuits conference, digest of technical papers, pp 366–367
145. Van den Bosch A, Borremans M, Steyaert M, Sansen W (2001) A 10-bit 1-GSample/s Nyquist current-steering CMOS D/A converter. IEEE J Solid-State Circuits 36:315–324
146. Miki T, Nakamura Y, Nakaya M, Asai S, Akasaka Y, Horiba Y (1986) An 80-MHz 8-bit CMOS D/A converter. IEEE J Solid-State Circuits 21:983–988
147. Bastiaansen CAA, Groeneveld DWJ, Schouwenaars HJ, Termeer HAH (1991) A 10-b 40-MHz 0.8-μm CMOS current-output D/A converter. IEEE Journal Solid-State Circuits 26:917–921
148. Doris K, Briaire J, Leenaerts D, Vertregt M, van Roermund A (2005) A 12b 500MS/s DAC with >70 dB SFDR up to 120 MHz in 0.18 μm CMOS. In: IEEE international solid-state circuits conference, digest of technical papers, pp 116–588
149. Van der Plas GAM, Vandenbussche J, Sansen W, Steyaert MSJ, Gielen GGE (1999) A 14-bit intrinsic accuracy Q^2 random walk CMOS DAC. IEEE J Solid-State Circuits 34:1708–1718
150. Jewett B, Liu J, Poulton K (2005) A 1.2GS/s 15b DAC for precision signal generation. In: IEEE international solid-state circuits conference, digest of technical papers, pp 110–111
151. Chen T, Gielen GGE (2007) A 14-bit 200-MHz current-steering DAC with switching-sequence post-adjustment calibration. IEEE J. Solid-State Circuits 42:2386–2394
152. Lin C-H, van der Goes FML, Westra JR, Mulder J, Lin Y, Arslan E, Ayranci E, Liu X, Bult K (2009) A 12 bit 2.9 GS/s DAC With IM3 <-60 dBc beyond 1 GHz in 65 nm CMOS. IEEE J Solid-State Circuits 44:3285–3293
153. Mercer DA (2007) Low-power approaches to high-speed current-steering digital-to-analog converters in 0.18-μm CMOS. IEEE J Solid-State Circuits 42:1688–1698
154. Park S, Kim G, Park SC, Kim W (2002) A digital-to-analog converter based on differential-quad switching. IEEE J Solid-State Circuits 37:1335–1338
155. Engel G, Kuo S, Rose S (2012) A 14-b 3/6GS/s current-steering RF DAC in 0.18-μm CMOS with 66 dB ACLR at 2.9 GHz. In: IEEE international solid-state circuits conference, digest of technical papers, pp 458–459
156. van de Vel H, Briaire J, Bastiaansen C, van Beek P, Geelen G, Gunnink H, Jin Y, Kaba M, Luo K, Paulus E, Pham B, Relyveld W, Zijlstra P (2014) A 240 mW 16b 3.2GS/s DAC in 65nm CMOS with <-80dBc IM3 up to 600 MHz. In: IEEE international solid-state circuits conference, digest of technical papers, pp 206–207

157. Schafferer B, Adams R (2004) A 3V CMOS 400 mW 14b 1.4GS/s DAC for multi-carrier applications. In: IEEE international solid-state circuits conference, digest of technical papers, pp 360–361
158. Schofield W, Mercer D, Onge LS (2003) A 16b 400MS/s DAC with < -80dBc IMD to 300MHz and < -160dBm/Hz noise power spectral density. In: IEEE international solid-state circuits conference, digest of technical papers, pp 126–127
159. Engel G, Clara M, Zhu H, Wilkins P (2015) A 16-bit 10Gsps current steering RF DAC in 65 nm CMOS achieving 65dBc ACLR multi-carrier performance at 4.5 GHz Fout. In: Symposium on VLSI circuits digest of technical papers, pp C166–167
160. Su DK, Wooley BA (1993) A CMOS oversampling D/A converter with a current-mode semidigital reconstruction filter. IEEE J Solid-State Circuits 28:1224–1233
161. Barkin DB, Lin ACY, Su DK, Wooley BA (2004) A CMOS oversampling bandpass cascaded D/A converter with digital FIR and current-mode semi-digital filtering. IEEE J Solid-State Circuits 39:585–593
162. Suarez RE, Gray PR, Hodges DA (1975) All-MOS charge-redistribution analog-to-digital conversion techniques. II. IEEE J Solid-State Circuits 10:379–385
163. Song B-S, Lee S-H, Tompsett MFA (1990) 10-b 15-MHz CMOS recycling two-step A/D converter. IEEE J Solid-State Circuits 25:1328–1338
164. Liu C-C, Chnag S-J, Huang G-Y, Lin Y-Z (2010) A 10-bit 50-MS/s SAR ADC with a monotonic capacitor switching procedure. IEEE J Solid-State Circuits 45:731–740
165. Lim Y, Flynn MP (2015) A 1 mW 71.5 dB SNDR 50 MS/s 13 bit fully differential ring amplifier based SAR-assisted pipeline ADC. IEEE J Solid-State Circuits 50:2901–2911
166. Pelgrom MJM, Roorda M (1988) An algorithmic 15 bit CMOS digital-to-analog converter. IEEE J Solid-State Circuits 23:1402–1405
167. Matsumoto H, Watanabe K (1986) Switched-capacitor algorithmic digital-to-analog converters. IEEE Trans Circuits Syst 33:721–724
168. Philips K, van den Homberg J, Dijkmans C (1999) PowerDAC: a single-chip audio DAC with a 70%-efficient power stage in 0.5 μm CMOS. In: IEEE international solid-state circuits conference, digest of technical papers, pp 154–155
169. Fan Q, Huijsing JH, Makinwa KAA (2012) A 21 nV/$\sqrt{\text{Hz}}$ Hz chopper-stabilized multi-path current-feedback instrumentation amplifier with 2 μV offset. IEEE J Solid-State Circuits 47:464–475
170. Blanken PG, Menten SEJ (2002) A 10 μV-offset 8 kHz bandwidth 4th-order chopped $\Sigma\Delta$ A/D converter for battery management. In: IEEE international solid-state circuits conference, digest of technical papers, pp 388–389
171. Carley L (1989)A noise-shaping coder topology for 15+ bit converters. IEEE J Solid-State Circuits 24:267–273
172. Nys OJAP, Henderson RK (1996) An analysis of dynamic element matching techniques in sigma-delta modulation. In: IEEE international symposium on circuits and systems, pp 231–234
173. Henderson RK, Nys OJAP (1996) Dynamic element matching techniques with arbitrary noise shaping function. In: IEEE international symposium on circuits and systems, pp 293–296
174. Story MJ (1992) Digital to analog converter adapted to select input sources based on a preselected algorithm once per cycle of a sampling signal. US patent 5-138-317
175. Maloberti A (1990/1991) Convertitore Digitale Analogico Sigma-Delta Multilivello con Matching Dinamico degli Elementi, Tesi di Laurea, Universita degli Studi di Pavia (this thesis was not available)
176. Miller MR, Petrie CS (2003) A multibit sigma-delta ADC for multimode receivers. IEEE J Solid-State Circuits 38:475–482
177. Risbo L, Hezar R, Kelleci B, Kiper H, Fares M (2011) A 108 dB DR. 120 dB THD and 0.5V_{rms} output audio DAC with inter-symbol-interference-shaping algorithm in 45 nm CMOS. In: IEEE international solid-state circuits conference, digest of technical papers, pp 484–485

178. Chiang S-HW, Sun H, Razavi B (2014) A 10-bit 800-MHz 19-mW CMOS ADC. IEEE J Solid-State Circuits 49:935–949
179. van der Goes F, Ward CM, Astgimath S, Yan H, Riley J, Zeng Z, Mulder J, Wang S, Bult K (2014) A 1.5 mW 68 dB SNDR 80 Ms/s 2 interleaved pipelined SAR ADC in 28 nm CMOS. IEEE J Solid-State Circuits 49:2835–2845
180. Fiedler HL, Hoefflinger B, Demmer W, Draheim P (1981) A 5-bit building block for 20 MHz A/D converters. IEEE J Solid-State Circuits 26:151–155
181. Nauta B, Venes AGW (1995) A 70 Ms/s 110 mW 8-b CMOS folding and interpolating A/D converter. IEEE J Solid-State Circuits 30:1302–1308
182. Yin G, Op't Eynde F, Sansen W (1992) A high-speed CMOS comparator with 8-b resolution. IEEE J Solid-State Circuits 27:208–211
183. Venes AGW, van de Plassche RJ (1996) An 80-MHz, 80-mW, 8-b CMOS folding A/D converter with distributed track-and-hold preprocessing. IEEE J Solid-State Circuits 31:1846–1853
184. Ellersick W, Chih-Kong KY, Horowitz M, Dally W (1999) GAD: a 12-GS/s CMOS 4-bit A/D converter for an equalized multi-level link. In: Symposium on VLSI circuits, digest of technical papers, pp 49–52
185. Kobayashi T, Nogami K, Shirotori T, Fujimoto Y (1993) A current-controlled latch sense amplifier and a static power-saving input buffer for low-power architecture. IEEE J Solid-State Circuits 28:523–527
186. Montanaro J et al (1996) A 160 MHz, 32b, 0.5W CMOS RISC microprocessor. IEEE J Solid-State Circuits 31:1703–1714
187. Verbruggen B, Craninckx J, Kuijk M, Wambacq P, Van der Plas G (2008) A 2.2 mW 5b 1.75GS/s folding flash ADC in 90 nm digital CMOS. In: IEEE international solid-state circuits conference, digest of technical papers, pp 252–611
188. Schinkel D, Mensink E, Klumperink E, van Tuijl E, Nauta B (2007) A double-tail latch-type voltage sense amplifier with 18ps Setup+Hold Time. In: IEEE international solid-state circuits conference, digest of technical papers, pp 314–315
189. Wu J-T, Wooley BA (1988) A 100-MHz pipelined CMOS comparator. IEEE J Solid-State Circuits 23:1379–1385
190. Fukushima N, Yamada T, Kumazawa N, Hasegawa Y, Soneda M (1989) A CMOS 40 MHz 8b 105 mW two-step ADC. In: International solid-state circuits conference, digest of technical papers, pp 14–15
191. Atherton JH, Simmonds HT (1992) An offset reduction technique for use with CMOS integrated comparators and amplifiers. IEEE J Solid-State Circuits 27:1168–1175
192. Haas M, Draxelmayr D, Kuttner F, Zojer B (1990) A monolithic triple 8-bit CMOS video coder. IEEE Trans Consumer Electron 36:722–729
193. Wong K-LJ, Yang C-KK (2004) Offset compensation in comparators with minimum input-referred supply noise. IEEE J Solid-State Circuits 37:837–840
194. Chen VH-C, Pileggi L (2014) A 69.5 mW 20 GS/s 6b time-interleaved ADC with embedded time-to-digital calibration in 32 nm CMOS SOI. IEEE J Solid-State Circuits 49:2891–2901
195. Schvan P, Pollex D, Wang S-C, Falt C, Ben-Hamida N (2006) A 22GS/s 5b ADC in 0.13μm SiGe BiCMOS. In: International solid-state circuits conference, digest of technical papers, pp 572–573
196. Reyhani H, Quinlan P (1994) A 5 V, 6-b, 80 Ms/s BiCMOS flash ADC. IEEE J Solid-State Circuits 29:873–878
197. Vorenkamp P, Verdaasdonk JPM (1992) A 10b 50MHz pipelined ADC. In: International solid-state circuits conference, digest of technical papers, pp 32–33
198. Kattmann K, Barrow J (1991) A technique for reducing differential nonlinearity errors in flash A/D converters. In: International solid-state circuits conference, digest of technical papers, pp 170–171
199. Scholtens PCS, Vertregt M (2002) A 6-b 1.6-Gsample/s flash ADC in 0.18μm CMOS using averaging termination. IEEE J Solid-State Circuits 37:1599–1609

200. Bult K, Buchwald A (1997) An embedded 240-mW 10-b 50-MS/s CMOS ADC in 1-mm^2. IEEE J Solid-State Circuits 32:1887–1895
201. Uyttenhove K, Vandenbussche J, Lauwers E, Gielen GGE, Steyaert MSJ (2003) Design techniques and implementation of an 8-bit 200-MS/s interpolating/averaging CMOS A/D converter. IEEE J Solid-State Circuits 38:483–494
202. Van De Grift REJ, Rutten IWJM, van der Veen M (1987) An 8-bit video ADC incorporating folding and interpolation techniques. IEEE J Solid-State Circuits 22:944–953
203. Vorenkamp P, Roovers R (1997) A 12-b, 60-MS/s cascaded folding and interpolating ADC. IEEE J Solid-State Circuits 32:1876–1886
204. Van De Plassche RJ, van der Grift REJ (1979) A high-speed 7 bit A/D converter. IEEE J Solid-State Circuits 14:938–943
205. Hoogzaad G, Roovers R (1999) A 65-mW, 10-bit, 40-MS/s BiCMOS Nyquist ADC in 0.8 mm^2. IEEE J Solid-State Circuits 34:1796–1802
206. Choe MJ, Song B-S, Bacrania K (2000) A 13b 40MS/s CMOS pipelined folding ADC with background offset trimming. In: International solid-state circuits conference, digest of technical papers, pp 36–37
207. Pernillo J, Flynn MP (2011) A 1.5-GS/s flash ADC with 57.7-dB SFDR and 6.4-Bit ENOB in 90 nm digital CMOS. IEEE Trans Circuits Syst II Exp Briefs 58:837–841
208. Shu Y-S (2012) A 6b 3GS/s 11 mW fully dynamic flash ADC in 40nm CMOS with reduced number of comparators. In: Symposium on VLSI circuits digest of technical papers, pp 26–27
209. van der Ploeg H, Vertregt M, Lammers M (2006) A 15-bit 30-MS/s 145-mW three-step ADC for imaging applications. IEEE J Solid-State Circuits 41:1572–1577
210. Lewis SH, Gray PR (1987) A pipelined 5-MS/s 9-bit analog-to-digital converter. IEEE J Solid-State Circuits 22:954–961
211. Shimizu Y, Murayama S, Kudoh K, Yatsuda H, Ogawa A (2006) A 30mW 12b 40MS/s subranging ADC with a high-gain offset-canceling positive-feedback amplifier in 90nm digital CMOS. In: International solid-state circuits conference, digest of technical papers, pp 802–803
212. Kusumoto K, Matsuzawa A, Murata K (1993) A 10-b 20-MHz 30-mW pipelined interpolating CMOS ADC. IEEE J Solid-State Circuits 28:1200–1206
213. Mulder J et al (2004) A 21 mW 8b 125MS/s ADC occupying 0.09 mm^2 in 0.13 μm CMOS. In: IEEE international solid-state circuits conference, digest of technical papers, pp 260–261
214. Lee CC, Flynn MP (2011) A SAR-assisted two-stage pipeline ADC. IEEE J Solid-State Circuits 46:859–869
215. Moreland C, Murden F, Elliott M, Young J, Hensley M, Stop R (2000) A 14b 100MS/s subranging ADC. IEEE J Solid-State Circuits 35:1791–1798
216. McCharles R, Hodges D (1978) Charge circuits for analog LSI. IEEE Trans Circuits Systems 25:490–497
217. Cho TB, Gray PR (1995) A 10 b, 20 Msample/s, 35 mW pipeline A/D converter. IEEE J Solid-State Circuits 30:166–172
218. Singer LA, Brooks TL (1996) A 14-bit 10-MHz calibration-free CMOS pipelined A/D converter. In: Proceedings symposium on VLSI circuits, pp 94–95
219. Ali AMA, Dillon C, Sneed R, Morgan AS, Bardsley S, Kornblum J, Wu L (2006) A 14-bit 125MS/s IF/RF sampling pipelined ADC with 100 dB SFDR and 50 fs jitter. IEEE J Solid-State Circuits 41:1846–1855
220. Li PW, Chin MJ, Gray PR, Castello R (1984) A ratio-independent algorithmic analog-to-digital conversion technique. IEEE J Solid-State Circuits 19:828–836
221. Chiu Y, Gray PR, Nikolic B (2004) A 14-b 12-MS/s CMOS pipeline ADC with over 100-dB SFDR. IEEE J Solid-State Circuits 39:2139–2151
222. Wang H, Wang X, Hurst PJ, Lewis SH (2009) Nested digital background calibration of a 12-bit pipelined ADC without an input SHA. IEEE J Solid-State Circuits 44:2780–2789
223. Bult K, Geelen GJGM (1990) A fast-settling CMOS op amp for SC circuits with 90-dB DC gain. IEEE J Solid-State Circuits 25:1379–1384

224. Chai Y, Wu J-T (2012) A 5.37mW 10b 200MS/s dual-path pipelined ADC. In: International solid-state circuits conference, digest of technical papers, pp 462–463
225. Wang X, Hurst PJ, Lewis SH (2004) A 12-bit 20-MS/s pipelined analog-to-digital converter with nested digital background calibration. IEEE J Solid-State Circuits 39:1799–1808
226. Murmann B (2012) Thermal noise in track-and-hold circuits: analysis and simulation techniques. IEEE Solid-State Circuits Mag 4:46–54
227. Karanicolas AN, Lee H-S, Barcrania KL (1993) A 15-b 1-MS/s digitally self-calibrated pipeline ADC. IEEE J Solid-State Circuits 28:1207–1215
228. Nagaraj K, Fetterman HS, Anidjar J, Lewis SH, Renninger RG (1997) A 250-mW, 8-b, 52-MSs/s parallel-pipelined A/D converter with reduced number of amplifiers. IEEE J Solid-State Circuits 32:312–320
229. Murmann B, Boser BE (2003) A 12-bit 75-MS/s pipelined ADC using open-loop residue amplification. IEEE J Solid-State Circuits 38:2040–2050
230. Iroaga E, Murmann B (2007) A 12-Bit 75-MS/s pipelined ADC using incomplete settling. IEEE J Solid-State Circuits 42:748–756
231. Geelen G, Paulus E, Simanjuntak D, Pastoor H, Verlinden R (2006) A 90nm CMOS 1.2V 10b power and speed programmable pipelined ADC with 0.5pJ/conversion-step. In: International solid-state circuits conference, digest of technical papers, pp 214–215
232. Bardsley S, Dillon C, Kummaraguntla R, Lane C, Ali AMA, Rigsbee B, Combs D (2006) A 100-dB SFDR 80-MSPS 14-bit 0.35-μm BiCMOS pipeline ADC. IEEE J Solid-State Circuits 41:2144–2153
233. Lee BG, Min BM, Manganaro G, Valvano JW (2008) A 14b 100 MS/s pipelined ADC with a merged active S/H and first MDAC. In: International solid-state circuits conference, digest of technical papers, pp 248–249
234. van de Vel H, Buter B, van der Ploeg H, Vertregt M, Geelen G, Paulus E (2009) A 1.2 V 250-mW 14-b 100 MS/s digitally calibrated pipeline ADC in 90-nm CMOS. IEEE J Solid-State Circuits 44:1047–1056
235. Chen C-Y, Wu J, Hung J-J, Li T, Liu W, Shih W-T (2012) A 12-bit 3 GS/s pipeline ADC with 0.4 mm^2 and 500 mW in 40 nm digital CMOS. IEEE J Solid-State Circuits 47:1013–1021
236. Tseng C-J, Chen H-W, Shen W-T, Cheng W-C, Chen H-S (2012) A 10-b 320-MS/s stage-gain-error self-calibration pipeline ADC. IEEE J Solid-State Circuits 47:1334–1343
237. Ali AMA et al (2014) A 14 Bit 1 GS/s RF sampling pipelined ADC with background calibration. IEEE J Solid-State Circuits 49:2857–2867
238. Min BM, Kim P, Bowman FW, Boisvert DM, Aude AJ (2003) A 69-mW 10-bit 80-MS/s pipelined CMOS ADC. IEEE J Solid-State Circuits 38:2031–2039
239. Lee BG, Tsang RM (2009) A 10-bit 50 MS/s pipelined ADC with capacitor-sharing and variable-gm opamp. IEEE J Solid-State Circuits 44:883–890
240. Mehr I, Singer L (2000) A 55-mW 10-bit 40-MS/s Nyquist-rate CMOS ADC. IEEE J Solid-State Circuits 35:318-323
241. Sepke T, Fiorenza JK, Sodini CG, Holloway P, Lee H-S (2006) Comparator-based switched-capacitor circuits for scaled CMOS technologies. In: International solid-state circuits conference, digest of technical papers, pp 812–821
242. Brooks L, Lee H-S (2009) A 12b 50MS/s fully differential zero-crossing-based ADC without CMFB. In: International solid-state circuits conference, digest of technical papers, pp 166–167
243. McCreary JL, Gray PR (1975) All-MOS charge redistribution analog-to-digital conversion techniques I. IEEE J Solid-State Circuits 10:371-379
244. Hamade AR (1978) A single chip all-MOS 8-bit A/D converter. IEEE J Solid-State Circuits 13:785–791
245. Wei H et al (2012) An 8-b 400-MS/s 2-b-per-cycle SAR ADC with resistive DAC. IEEE J Solid-State Circuits 47:2763–2772
246. Craninckx J, Van der Plas G (2007) A 65fJ/conversion-step 0-to-50MS/s 0-to-0.7mW 9b charge-sharing SAR ADC in 90nm Digital CMOS. In: International solid-state circuits conference, digest of technical papers, pp 246-247

247. Agnes A, Bonizzoni E, Malcovati P, Maloberti F (2008) A 9.4-ENOB 1V 3.8μW 100kS/s SAR ADC with time-domain comparator. In: International solid-state circuits conference, digest of technical papers, pp 246–247
248. van Elzakker M, van Tuijl E, Geraedts P, Schinkel D, Klumperink E, Nauta B (2008) A 1.9μW 4.4fJ/conversion-step 10b 1MS/s charge-redistribution ADC. In: International solid-state circuits conference, digest of technical papers, pp 244–245
249. Ginsburg BP, Chandrakasan AP (2007) 500-MS/s 5-bit ADC in 65-nm CMOS with split capacitor array DAC. IEEE J Solid-State Circuits 42:739–747
250. Chen DG, Law MK, Lian Y, Bermak A (2016) Low-power CMOS laser doppler imaging using non-CDS pixel readout and 13.6-bit SAR ADC. IEEE Trans Biomed Circuits Syst 10:186–199
251. Kuttner F (2002) A 1.2V 10 b 20 MS/s non-binary successive approximation ADC in 0.13μm CMOS. In: International solid-state circuits conference, digest of technical papers, pp 176–177
252. Harpe PJA, Zhou C, Bi Y, van der Meijs NP, Wang X, Philips KJP, Dolmans G, de Groot HWH (2011) A 26 μW 8 bit 10 MS/s asynchronous SAR ADC for low energy radios. IEEE J Solid-State Circuits 46:1585–1595
253. Harpe PJA, Cantatore E, van Roermund AHM (2014) An oversampled 12/14b SAR ADC with noise reduction and linearity enhancements achieving up to 79.1 dB SNDR. In: International solid-state circuits conference, digest of technical papers, pp 194–195
254. Liu W, Chang y, Hsien S-K et al (2009) A 600 mW 30 mW 0.13 μm CMOS ADC array achieving over 60dB SFDR with adaptive digital equalization. In: International solid-state circuits conference, digest of technical papers, pp 82-83
255. Bannon A, Hurrell CP, Hummerston D, Lyden C (2014) An 18-b 5-MS/s SAR ADC with 100.2 dB dynamic range. In: Symposium on VLSI circuits digest of technical papers, pp 1–2
256. Hesener M, Ficher T, Hanneberg A, Herbison D, Kuttner F, Wenskel H (2007) A 14b 40MS/s redundant SAR ADC with 480 MHz in 0.13 pm CMOS. In: International solid-state circuits conference, digest of technical papers, pp 248–249
257. Harpe P, Zhang Y, Dolmans G, Philips K, de Groot H (2012) A 7-to-10b, 0-to-4MS/s flexible SAR ADC with 6.5-to-16fJ/conversion-step. In: International solid-state circuits conference, digest of technical papers, pp 472–473
258. Pelgrom MJM, van Rens AC, Vertregt M, Dijkstra MB (1994) A 25-Ms/s 8-bit CMOS A/D converter for embedded application. IEEE J Solid-State Circuits 29:879–886
259. Murray B, Menting H (1992) A highly integrated D2MAC decoder. In: IEEE international conference on consumer electronics, digest of technical papers, pp 56–57
260. Shih C, Gray PR (1986) Reference refreshing cyclic analog-to-digital and digital-to-analog converters. IEEE J Solid-State Circuits 21:544–554
261. Ginetti B, Jespers P, Vandemeulebroecke A (1992) A CMOS 13-b cyclic. A/D converter. IEEE J Solid-State Circuits 27:957–964
262. Mase M, Kawahito S, Sasaki M, Wakamori Y, Furuta M (2005) A wide dynamic range CMOS image sensor with multiple exposure-time signal outputs and 12-bit column-parallel cyclic A/D converters. IEEE J Solid-State Circuits 40:2787–2795
263. Lim S et al (2011) A 240-frames/s 2.1-Mpixel CMOS image sensor with column-shared cyclic ADCs. IEEE J Solid-State Circuits 46:2073-2083
264. Naraghi S, Courcy M, Flynn MP (2009) A 9b 14 μW 0.06 mm^2 PPM ADC in 90nm digital CMOS. In: IEEE International solid-state circuits conference digest of technical papers, pp 168–169
265. Snoeij MF, Theuwissen AJP, Makinwa KAA, Huijsing JH (2007) Multiple-ramp column-parallel ADC architectures for CMOS image sensors. IEEE J Solid-State Circuits 42:2986–2977
266. Howard BK (1955) Binary quantizer. US patent 2-715-678
267. van der Ploeg H, Hoogzaad G, Termeer HAH, Vertregt M, Roovers RLJ (2001) A 2.5V, 12b, 54MS/s, 0.25um CMOS ADC. In: International solid-state circuits conference, digest of technical papers, pp 132–133

268. Mark JW, Todd TD (1981) A nonuniform sampling approach to data compression. IEEE Trans Commun 29:24–32
269. Allier E, Goulier J, Sicard G, Dezzani A, Andre E, Renaudin M (2005) A 120-nm low power asynchronous ADC. In: International symposium on low-power electronics and design, pp 60–65
270. Trakimas M, Sonkusale SR (2011) An adaptive resolution asynchronous ADC architecture for data compression in energy constrained sensing applications. IEEE Trans Circuits Syst I 58:921–934
271. Lin C-S, Liu B-D (2003) A new successive approximation architecture for low-power low-cost CMOS A/D converter. IEEE J Solid-State Circuits 38:54–62
272. Chen S-WM, Brodersen RW (2006) A 6-bit 600-MS/s 5.3-mW asynchronous ADC in 0.13-μm CMOS. IEEE J Solid-State Circuits 41:2669–2680
273. Straayer MZ, Perrott MH (2008) A 12-Bit, 10-MHz bandwidth, continuous-time ADC with a 5-Bit, 950-MS/s VCO-based quantizer. IEEE J Solid-State Circuits 43:805–814
274. Jansson J-P, Mantyniemi A, Kostamovaara J (2006) A CMOS time-to-digital converter with better than 10 ps single-shot precision. IEEE J Solid-State Circuits 41:1286–1296
275. Chen P, Liu S-L, Wu J (2000) A CMOS pulse-shrinking delay element for time interval measurement. IEEE Trans Circuits Syst 47:954–958
276. Rahkonen TE, Kostamovaara JT (1993) The use of stabilized CMOS delay lines for the digitization of short time intervals. IEEE J Solid-State Circuits 28:887–894
277. van der Ploeg H (1997) The nonius analog-to-digital converter. Internal Philips Research report, University Twente B.Sc. report, Supervisor M. Pelgrom
278. Groza VZ (2001) High-resolution floating-point ADC. IEEE Trans Instrum Meas 50:1812–1829

9. Time-Interleaving

279. Black WC, Hodges DA (1980) Time-interleaved converter arrays. IEEE J Solid-State Circuits 15(12):1022–1029
280. Pelgrom MJM, Jochijms A, Heijns H (1987) A CCD delay line for video applications. IEEE Trans Consum Electron 33:603–609
281. Kapusta R et al (2013) A 14-b 80 MS/s SAR ADC with 73.6 dB SNDR in 65 nm CMOS. IEEE J Solid-State Circuits 48:3059–3066
282. El-Chammas M, Murmann B (2011) A 12-GS/s 81-mW 5-bit time-interleaved flash ADC with background timing skew calibration. IEEE J Solid-State Circuits 46(4):838–847
283. Doris K, Janssen E, Nani C, Zanikopoulos A, Van der Wiede G (2011) A 480 mW 2.6 GS/s 10b Time-Interleaved ADC With 48.5 dB SNDR up to Nyquist in 65 nm CMOS. IEEE J Solid-State Circuits 46:2821–2833
284. Kull L et al (2014) A 90-GS/s 8-b 667 mW 64x Interleaved SAR ADC in 32 nm Digital SOI CMOS. In: International solid-state circuits conference, digest of technical papers, pp 378–379
285. Kull L, Pliva J, Toifl T, Schmatz M, Francese PA, Menolfi C, Brändli M, Kossel M, Morf T, Meyer Andersen T, Leblebici Y (2016) Implementation of low-power 6–8 b 30–90 GS/s time-interleaved ADCs with optimized input bandwidth in 32 nm CMOS. IEEE J Solid-State Circuits 51:636–648
286. Louwsma SM et al (2008) A 1.35 GS/s, 10 b, 175 mW time-interleaved AD converter in 0.13μm CMOS. IEEE J Solid-State Circuits 43:778–786
287. Varzaghani A et al (2013) A 10.3-GS/s, 6-bit flash ADC for 10G ethernet applications. IEEE J Solid-State Circuits 48(8):3038–3048
288. Hsu C-C, Huang F-C, Shih C-Y, Huang C-C, Lin Y-H, Lee C-C, Razavi B (2007) An 11b 800MS/s time-interleaved ADC with digital background calibration. In: International solid-state circuits conference, digest of technical papers, pp 164–165
289. Razavi B (2013) Design considerations for interleaved ADCs. IEEE J Solid-State Circuits 48(8):1806–1817
290. Janssen E et al (2013) An 11b 3.6GS/s Time-Interleaved SAR ADC in 65nm CMOS. In: International solid-state circuits conference, digest of technical papers, pp 464–465

291. Straayer M et al (2016) A 4GS/s time-interleaved RF ADC in 65nm CMOS with 4GHz input bandwidth. In: International solid-state circuits conference, digest of technical papers, pp 464–465
292. Ram'rez-Angulo J, Carvajal RG, Torralba A, Galán J, VegaLeal AP, Tombs J (2002) The flipped voltage follower: a useful cell for low-voltage low power circuit design. In: Proceedings ISCAS 2002, vol 3, pp 615–618
293. Kurosawa N, Kobayashi H, Maruyama K, Sugawara H, Kobayashi K (2001) Explicit analysis of channel mismatch effects in time-interleaved ADC systems. IEEE Trans Circuits Syst I: Fundam Theory Appl 48:261–271
294. Olieman E, Annema A-J, Nauta B (2015) An interleaved full Nyquist high-speed DAC technique. IEEE J Solid-State Circuits, 50:704–715
295. Olieman E (2016) Time-interleaved high-speed D/A converters. Ph.D. thesis, University Twente
296. Wu J et al (2013) A 5.4GS/s 12b 500mW pipeline ADC in 28nm CMOS. In: Symposium on VLSI circuits digest of technical papers, pp C92–93
297. Verbruggen B, Iriguchi M, Craninckx J (2012) A 1.7mW 11b 250MS/s 2× interleaved fully dynamic pipelined SAR ADC in 40nm digital CMOS. In: International solid-state circuits conference, digest of technical papers, pp 466–467
298. Brandolini M et al (2015) A 5 GS/s 150 mW 10 b SHA-less pipelined/SAR hybrid ADC for direct-sampling systems in 28 nm CMOS. IEEE J Solid-State Circuits 50:2922–2934
299. Verma S, Kasapi A, Lee L, Liu D (2013) A 10.3 GS/s 6b flash ADC for 10G ethernet applications. In: International solid-state circuits conference, digest of technical papers, pp 462–463
300. Stepanovic D, Nikolic B (2013) A 2.8 GS/s 44.6 mW time-interleaved ADC achieving 50.9 dB SNDR and 3 dB effective resolution bandwidth of 1.5 GHz in 65 nm CMOS. IEEE J Solid-State Circuits 48:971–982
301. Setterberg B et al (2013) A 14 b 2.5 GS/s 8-way-interleaved pipelined ADC with background calibration and digital dynamic linearity correction. In: International solid-state circuits conference, digest of technical papers, pp 466–467

10. Sigma-Delta Conversion

302. de Jager F (1952) Delta modulation, a method of PCM transmission using the 1-unit code. Philips Res Rep 7:442–466
303. Cutler C (1960) Transmission systems employing quantization. U.S. Patent 2-927-962
304. Widrow B (1956) A study of rough amplitude quantization by means of Nyquist sampling theory. IRE Trans Circuit Theory CT-3:266–276
305. Inose H, Yasuda Y, Murakami J (1962) A telemetering system by code modulation- $\Delta\Sigma$ modulation. IRE Trans Space Electron Telem SET-8:204–209; or Proc IEEE 51:1524–1535 (1963)
306. Candy JC, Temes GC (eds) Oversampling delta-sigma data converters: theory, design and simulation IEEE, New York (1992)
307. Norsworthy SR, Schreier R, Temes GC (eds) (1997) Delta-sigma data converters: theory, design, and simulation. IEEE Press, Piscataway. ISBN:0-7803-1045–4
308. Schreier R, Temes GC (2004) Understanding delta-sigma data converters. Wiley, New York. ISBN:0-471-46585–2
309. de la Rosa JM (2011) Sigma-delta modulators: tutorial overview, design guide, and state-of-the-art survey. IEEE Trans Circuits Syst I 58:1–21
310. de la Rosa JM, Schreier R, Pun K-P, Pavan S (2015) Next-generation delta-sigma converters: trends and perspectives. IEEE J Emerging Sel Top Circuits Syst 5:484–499
311. Naus PJA, Dijkmans EC (1988) Low signal level distortion in sigma-delta modulators. In: 84th convention of the audio engineering society, Paris
312. Bonizzoni E, Pena Perez A, Maloberti F, Garcia-Andrade M (2008) Third-order sigma-delta modulator with 61-dB SNR and 6-MHz bandwidth consuming 6 mW. In: European solid-state circuits conference, pp 218–221

313. Pena-Perez A, Bonizzoni E, Maloberti F (2012) A 88-dB DR, 84-dB SNDR very low-power single op-amp third-order modulator. IEEE J Solid-State Circuits 47:2107–2118
314. Gambini S, Rabaey J (2007) A 100-kS/s 65-dB DR sigma-delta ADC with 0.65V supply voltage. In: 33th European solid state circuits conference, pp 202–205
315. Vleugels K, Rabii S, Wooley BA (2001) A 2.5 v sigma-delta modulator for broadband communications applications. IEEE J Solid-State Circuits 36:1887–1889
316. Chae Y, Han G (2009) Low voltage, low power, inverter-based switched-capacitor delta-sigma modulator. IEEE J Solid-State Circuits 44:369–373
317. Hart A, Voinigescu SP (2009) A 1 GHz bandwidth low-pass sigma-delta ADC with 20–50 GHz adjustable sampling rate. IEEE J Solid-State Circuits 44:1401–1414
318. Lee WL, Sodini CG (1987) A topology for higher order interpolative coders. In: Proceedings of the IEEE international symposium on circuits and systems, pp 459–462
319. Chao KC-H, Nadeem S, Lee WL, Sodini CG (1990) A higher order topology for interpolative modulators for oversampling A/D converters. IEEE Trans Circuits Syst 37:309–318
320. Williams III LA, Wooley BA (1994) A third-order sigma-delta modulator with extended dynamic range. IEEE J Solid-State Circuits 29:193–202
321. Marabelli S, Fornasari A, Malcovati P, Maloberti F (2005) An 80-MHz, 14-bit bandpass mash sigma-delta/pipeline A/D converter with 5 MHz bandwidth for third generation mobile communication systems. Measurement 37:320–327
322. Williams L, Wooley B (1991) Third-order cascaded sigma-delta modulators. IEEE Trans Circuits Syst 38:489–498
323. Christen T, Burger T, Huang Q (2007) A 0.13 μm CMOS EDGE/UMTS/WLAN tri-mode ADC with-92dB THD. In: International solid-state circuits conference, digest of technical papers, pp 240–241
324. Breems LJ, Rutten R, van Veldhoven RHM, van der Weide G (2007) A 56 mW continuous-time quadrature cascaded $\Sigma\Delta$ modulator with 77 dB DR in a near zero-IF 20 MHz band. IEEE J Solid-State Circuits 42:2696–2705
325. Kulchycki SD, Trofin R, Vleugels K, Wooley BA (2008) A 77-dB dynamic range, 7.5-MHz hybrid continuous-time/discrete-time cascaded sigma delta modulator. IEEE J Solid-State Circuits 43:796–804
326. Gharbiya A, Johns DA (2009) A 12-bit 3.125 MHz bandwidth 0–3 MASH delta-sigma modulator. IEEE J Solid-State Circuits 44:2010–2018
327. Chae YC, Souri K, Makinwa KAA (2013) A 6.3 μW 20bit incremental zoom-ADC with 6 ppm INL and 1 μV offset. IEEE J Solid-State Circuits 48:3019–3027
328. Gönen B, Sebastiano F, van Veldhoven R, Makinwa KAA (2016) A 1.65 mW 0.16 mm^2 dynamic zoom-ADC with 107.5 dB DR in 20 kHz BW. In: International solid-state circuits conference, digest of technical papers, pp 282–283
329. Kup BMJ, Dijkmans EC, Naus PJA, Sneep J (1991) A bit-stream digital-to-analog converter with 18-b resolution. IEEE J Solid-State Circuits 26:1757–1763
330. Naus PJA, Dijkmans EC, Stikvoort EF, Holland DJ, Bradinal W (1987) A CMOS stereo 16-bit D/A converter for digital audio. IEEE J Solid-State Circuits 22:390–395
331. Adams R, Nguyen KQ (1998) A 113-dB SNR oversampling DAC with segmented noise-shaped scrambling. IEEE J Solid-State Circuits 33:1871–1878
332. Colonna V, Annovazzi M, Boarin G, Gandolfi G, Stefani F, Baschirotto A (2005) A 0.22-mm^2 7.25-mW per-channel audio stereo-DAC With 97-dB DR and 39-dB SNR. IEEE J Solid-State Circuits 40:1491–1498
333. Benabes P, Keramat M, Kielbasa R (1997) A methodology for designing continuous-time sigma-delta modulators. In: Proceedings European design and test conference, EDTC -97, pp 46–50
334. van der Zwan EJ, Dijkmans EC (1996) A 0.2 mW CMOS $\Sigma\Delta$ modulator for speech coding with 80 dB dynamic range. IEEE J Solid-State Circuits 31:1873–1880
335. Breems LJ, Dijkmans EC, Huijsing JH (2001) A quadrature data-dependent DEM algorithm to improve image rejection of a complex $\Sigma\Delta$ modulator. IEEE J Solid-State Circuits 36:1879–1886

336. Dong Y, Yang W, Schreier R et al (2014) A continuous-time 0–3 MASH ADC achieving 88 dB DR with 53 MHz BW in 28 nm CMOS. IEEE J Solid-State Circuits 49:2868–2877
337. Keller M, Buhmann A, Sauerbrey J, Ortmanns M, Manoli Y (2008) A comparative study on excess-loop-delay compensation techniques for continuous-time sigma-delta modulators. IEEE Trans Circuits Syst I 55:3480–3487
338. Bolatkale M (2013) High speed and wide bandwidth delta-sigma ADCs. PhD thesis, Technical University Delft. Available: http://repository.tudelft.nl
339. Bolatkale M, Breems LJ, Rutten R, Makinwa KAA (2011) A 4 GHz continuous-time ADC with 70 dB DR and 74 dBFS THD in 125 MHz BW. IEEE J Solid-State Circuits 46:2857–2868
340. Philips KJP (2005) $\Sigma\Delta$ AD conversion for signal conditioning. PhD thesis, Technical University Eindhoven
341. Philips K, Nuijten PACM, Roovers RLJ, van Roermund AHM, Chavero FM, Pallares MT, Torralba A (2004) A continuous-time $\Sigma\Delta$ ADC with increased immunity to interferers. IEEE J Solid-State Circuits 39:2170–2178
342. Nguyen K, Adams R, Sweetland K, Chen H (2005) A 106-dB SNR hybrid oversampling analog-to-digital converter for digital audio. IEEE J Solid-State Circuits, 40:2408–2415
343. Shettigar P, Pavan S (2012) A 15 mW 3.6GS/s CT-SD ADC with 36 MHz bandwidth and 83 dB DR in 90 nm CMOS. In: International solid-state circuits conference, digest of technical papers, pp 156–157
344. Shibata H, Schreier R, Yang W, Shaikh A, Paterson D, Caldwell T, Alldred D, Lai PW (2012) A DC-to-1GHz tunable RF SD ADC achieving DR = 74 dB and BW = 150 MHz at f0 = 450 MHz using 550 mW. In: IEEE international solid-state circuits conference, digest of technical papers, pp 150–151
345. Adams RW (1986) Design and implementation of an audio 18-bit analog-to-digital converter using oversampling techniques. J Audio Eng Soc 34:153–166
346. Naus PJA, Dijkmans EC (1991) Multi-bit oversampled $\Sigma\Delta$ A/D converters as front-end for CD players. IEEE J Solid-State Circuits 26:905–909
347. van Veldhoven RHM, Minnis BJ, Hegt HA, van Roermund AHM (2002) A 3.3-mW $\Sigma\Delta$ modulator for UMTS in 0.18-μm CMOS with 70-dB dynamic range in 2-MHz bandwidth. IEEE J Solid-State Circuits 37:1645–1652
348. Ranjbar M et al (2010) A 3.1 mW continuous-time $\Sigma\Delta$ modulator with 5-bit successive approximation quantizer for WCDMA. IEEE J Solid-State Circuits 45:1479–1491
349. Lipshitz SP, Vanderkooy J (2001) Why 1-bit sigma-delta conversion is unsuitable for high-quality applications, paper 5395. In: 110th convention of the audio engineering society, Amsterdam
350. Silva PGR, Breems LJ, Makinwa K, Roovers R, Huijsing JH (2007) An IF-to-baseband $\Sigma\Delta$ modulator for AM/FM/IBOC radio receivers with a 118 dB dynamic range. IEEE J Solid-State Circuits 42:1076–1089
351. Ouzounov SF, Roza E, Hegt JA, van de Weide G, van Roermund AHM (2006) Analysis and design of high-performance asynchronous sigma delta modulators with binary quantizer. IEEE J Solid-State Circuits 41:588–596
352. Vink J, van Rens J (1998) A CMOS multi-bit sigma-delta modulator for video applications. In: European solid-state circuits conference, pp 164–167
353. Park H, Nam KY, Su DK, Vleugels K, Wooley BA (2009) A 0.7-V 870-μW digital-audio CMOS sigma-delta modulator. IEEE J Solid-State Circuits 44:1078–1088
354. Engelen J, van de Plassche R, Stikoort E, Venes A (1999) A sixth-order continuous-time bandpass sigma-delta modulator for digital radio IF. IEEE J Solid-State Circuits 34:1753–1764
355. Schreier R, Abaskharoun N, Shibata H, Mehr I, Rose S, Paterson D (2006) A 375mW quadrature bandpass delta sigma ADC with 90dB DR and 8.5MHz BW at 44MHz. In: International solid-state circuits conference, digest of technical papers, pp 141–142
356. Ryckaert J et al (2009) A 2.4GHz low-power sixth-order RF bandpass SD converter in CMOS. IEEE J Solid-State Circuits 44:2873–2880

357. Harrison J, Nesselroth M, Mamuad R, Behzad A, Adams A, Avery S (2012) An LC bandpass SD ADC with 70dB SNDR over 20MHz bandwidth using CMOS DACs. In: International solid-state circuits conference, digest of technical papers, pp 146–147
358. Chae H, Jeong J, Manganaro G, Flynn M (2012) A 12mW low-power continuous-time bandpass SD modulator with 58dB SNDR and 24MHz bandwidth at 200 MHz IF. In: International solid-state circuits conference, digest of technical papers, pp 148–149
359. Martens E et al (2012) RF-to-baseband digitization in 40 nm CMOS with RF bandpass $\Delta\Sigma$ modulator and polyphase decimation filter. IEEE J Solid-State Circuits 47:990–1002
360. Tao H, Khoury JM (1999) A 400-Ms/s frequency translating bandpass sigma-delta modulator. IEEE J Solid-State Circuits 34:1741–1752
361. van Veldhoven RHM (2003) A triple-mode continuous-time $\Sigma\Delta$ modulator with switched-capacitor feedback DAC for a GSM-EDGE/ CDMA2000/UMTS receiver. IEEE J Solid-State Circuits 38:2069–2076
362. van der Zwan EJ, Philips K, Bastiaansen CAA (2000) A 10.7-MHz IF-to-baseband $\Sigma\Delta$ A/D conversion system for AM/FM radio receivers. IEEE J Solid-State Circuits 35:1810–1819
363. Bergveld HJ, van Kaam KMM, Leenaerts DMW, Philips KJP, Vaassen AWP, Wetzker G (2004) A low-power highly-digitized receiver for 2.4-GHz-band GFSK applications. In: IEEE radio frequency integrated circuits (RFIC) symposium, pp 347–350
364. Kavusi S, Kakavand H, El Gamal A (2006) On incremental sigma-delta modulation with optimal filtering. IEEE Trans Circuits Syst I: Fundam Theory Appl 53:1004–1015
365. Quiquempoix V, Deval P, Barreto A, Bellini G, Markus J, Silva J, Temes GC (2006) A low-power 22-bit incremental ADC. IEEE J Solid-State Circuits 41(7), 1562–1571
366. Oike Y, Gamal AE (2013) CMOS image sensor with per-column $\Sigma\Delta$ ADC and programmable compressed sensing. IEEE J Solid-State Circuits 48:318–328

11. Characterization and Specification

367. Irons FH (2000) The noise power ratio – theory and adc testing. IEEE Trans Instrum Meas 49:659–665
368. Milor LS (1998) A tutorial introduction to research on analog and mixed-signal circuit testing. IEEE Trans Circuits Syst II 45:1389–1407

Index

M. Pelgrom, *Analog-to-Digital Conversion*, DOI 10.1007/978-3-319-44971-5