CIRCUIT TECHNIQUES FOR LOW-VOLTAGE AND HIGH-SPEED A/D CONVERTERS

Mikko Waltari

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Electrical and Communications Engineering for public examination and debate in Auditorium S4 at Helsinki University of Technology (Espoo, Finland) on the 24th of June, 2002, at 12 o'clock noon.

Abstract

The increasing digitalization in all spheres of electronics applications, from telecommunications systems to consumer electronics appliances, requires analog-to-digital converters (ADCs) with a higher sampling rate, higher resolution, and lower power consumption. The evolution of integrated circuit technologies partially helps in meeting these requirements by providing faster devices and allowing for the realization of more complex functions in a given silicon area, but simultaneously it brings new challenges, the most important of which is the decreasing supply voltage.

Based on the switched capacitor (SC) technique, the pipelined architecture has most successfully exploited the features of CMOS technology in realizing high-speed high-resolution ADCs. An analysis of the effects of the supply voltage and technology scaling on SC circuits is carried out, and it shows that benefits can be expected at least for the next few technology generations. The operational amplifier is a central building block in SC circuits, and thus a comparison of the topologies and their low voltage capabilities is presented.

It is well-known that the SC technique in its standard form is not suitable for very low supply voltages, mainly because of insufficient switch control voltage. Two low-voltage modifications are investigated: switch bootstrapping and the switched opamp (SO) technique. Improved circuit structures are proposed for both. Two ADC prototypes using the SO technique are presented, while bootstrapped switches are utilized in three other prototypes.

An integral part of an ADC is the front-end sample-and-hold (S/H) circuit. At high signal frequencies its linearity is predominantly determined by the switches utilized. A review of S/H architectures is presented, and switch linearization by means of bootstrapping is studied and applied to two of the prototypes. Another important parameter is sampling clock jitter, which is analyzed and then minimized with carefully-designed clock generation and buffering.

The throughput of ADCs can be increased by using parallelism. This is demon-

strated on the circuit level with the double-sampling technique, which is applied to S/H circuits and a pipelined ADC. An analysis of nonidealities in double-sampling is presented. At the system level parallelism is utilized in a time-interleaved ADC. The mismatch of parallel signal paths produces errors, for the elimination of which a timing skew insensitive sampling circuit and a digital offset calibration are developed.

A total of seven prototypes are presented: two double-sampled S/H circuits, a time-interleaved ADC, an IF-sampling self-calibrated pipelined ADC, a current steering DAC with a deglitcher, and two pipelined ADCs employing the SO technique.

Preface

The greater becomes the volume of our sphere of knowledge, the greater also becomes its surface of contact with the unknown.

Jules Sagret

The above quote was printed in a yellow postcard—an invitation message from the student union—attached to the letter announcing my acceptance to the university in 1992. These ten years have really revealed me the wisdom of those words.

The research reported in this thesis has been carried out at Helsinki University of Technology, Electronic Circuit Design Laboratory between 1997 and 2002. During that time I have had a privilege of participating the Graduate School in Electronics, Telecommunications, and Automation (GETA) and working with research projects funded by Finnish National Technology Agency (TEKES) and Academy of Finland.

I want to express my gratitude to my supervisor professor Kari Halonen, who has given me the opportunity of working with these interesting projects. The fruits of this work wouldn't have been what they are without him pushing the design goals higher and encouraging to write and submit manuscripts to the topmost academic forums in this field.

I am specially thankful for Lauri Sumanen, who has been my partner in the two-man core team of Nyquist rate ADC research. His commitment and technical contribution to the projects we have carried out together has been invaluable. The greatest results of our cooperation are those two prototypes that we managed to publish in Journal of Solid-State Circuits and International Solid-State Circuits Conference. The junior team members in these projects, Jussi Pirkkalaniemi and Tuomas Korhonen, also deserve a mention.

My warmest thanks also go to Marko "Pikkis" Kosunen, with whom I shared an office for six years. Our daily discussions about technical and nontechnical matters is one of the things I will miss most from the laboratory.

From among my senior colleagues, some of which have already left the laboratory, I want specially mention Saska Lindfors, Jouko Vankka, and Kimmo Koli. Their help in figuring out the strengths and weaknesses of my new ideas as well as finding the relevant literature references has been irreplaceable. In addition to Pikkis and Saska two other old-timers, Mika "Läkä" Länsirinne and Rami Ahola, have also been great company to work and to have fun with.

Thanks for the exceptional atmosphere and team spirit in the laboratory belongs to all these people as well as those not mentioned by name. However, two persons whose impact has been especially strong are our secretary Helena Yllö and professor Veikko Porra.

The reviewers of this thesis, professors S. H. Lewis and R. J. van de Plassche, deserve my warmest thanks for their thorough acquaintance with the manuscript and their valuable comments and suggestions.

Last but not least I must thank the home front. The support from my wife Miia and her appreciation toward my work are those things that have helped me pushing on when the amount of work yet to be done has seemed unbearable. I also want to thank my parents Ritva and Eljas for the inspiring atmosphere in my childhood home, which awoke my interest in sciences and led my way to the university.

Espoo, May 2002.

Mikko Waltari

Contents

Symbols and Abbreviations

$\delta(t)$	Dirac's delta function
γ	Noise excess factor
μ	Carrier mobility
ω	Corner frequency
σ^2	Variance
A	Open-loop DC gain
A_{Vth}	Threshold voltage matching parameter
C_G	Gate capacitance
C_L	Load capacitance
C_{ox}	Gate oxide capacitance
f	Frequency
f_S	Sampling frequency
g_m	Transconductance
g_{ds}	Channel conductance
$h(t)$	Sampling function
I_D	Drain current
k	Boltzmann's constant
L	Channel length
L_{min}	Minimum channel length
N	Number of bits
R_{ON}	Switch on-resistance
T	Absolute temperature, sampling period

t	Time
T_{ox}	Oxide thickness
V_T	Threshold voltage
V_{BE}	Base-emitter voltage
V_{DD}	Positive supply voltage
V_{dsat}	Drain-source saturation voltage
V_{FS}	Full-scale voltage
V_{GS}	Gate-source voltage
V_{SS}	Negative supply voltage
V_{th}	Threshold voltage
W	Channel width
A/D	Analog-to-Digital
ADC	Analog-to-Digital Converter
BGR	BandGap Reference
BiCMOS	Bipolar CMOS
BJT	Bipolar Junction Transistor
CDS	Correlated Double Sampling
CFCS	Commutated Feedback Capacitor Switching
CM	Common Mode
CMFB	Common Mode FeedBack
CMOS	Complementary Metal Oxide Semiconductor
CP	Charge Pump
D/A	Digital-to-Analog
DAC	Digital-to-Analog Converter
dBc	Decibel relative to carrier
dBFS	Decibel relative to full-scale
DC	Direct Current
DFT	Discrete Fourier Transform
DLL	Delay Locked Loop

DNL	Differential Non-Linearity
DR	Dynamic Range
ENOB	Effective Number Of Bits
ERB	Effective Resolution Bandwidth
FPGA	Field Programmable Gate-Array
GaAs	Gallium Arsenide
GBW	Gain-bandwidth product
IC	Integrated Circuit
IF	Intermediate Frequency
INL	Integral Non-Linearity
LO	Local Oscillator
LSB	Least Significant Bit
MDAC	Multiplying Digital-to-Analog Converter
MOS	Metal Oxide Semiconductor
MOSFET	Metal Oxide Semiconductor Field Efect Transistor
MSB	Most Significant Bit
nMOS	n-channel Metal Oxide Semiconductor
OTA	Operational Transconductance Amplifier
PCB	Printed Circuit Board
PD	Phase Detector
PDA	Personal Digital Assistant
PLL	Phase Locked Loop
pMOS	p-channel Metal Oxide Semiconductor
Q	Quality factor
RHP	Right Half Plane
RSD	Redundant Signed Digit
S/H	Sample-and-Hold
SAW	Surface Acoustic Wave
SC	Switched Capacitor

SFDR	Spurious Free Dynamic Range
SiGe	Silicon Germanium
SINAD	see SNDR
$\text{sinc}(x)$	$\sin(x)/x$
SNDR	Signal-to-Noise-and-Distortion Ratio
SNR	Signal-to-Noise Ratio
SO	Switched Opamp
SoC	System on a Chip
SOI	Silicon On Insulator
T/H	Track-and-Hold
THD	Total Harmonic Distortion
VCO	Voltage Controlled Oscillator
VHDL	Very high speed integrated circuit Hardware Description Language
Vpp	Volts peak-to-peak

Chapter 1

Introduction

1.1 Motivation

For four decades the evolution of integrated circuits has followed Moore's law, according to which the number of transistors per square millimeter of silicon doubles every 18 months. At the same time transistors have become faster, making possible ever-increasing clock rates in digital circuits. This trend seems set to continue for at least another decade without slowing down. Thus, in the near future the processing power of digital circuits will continue to increase at an accelerating pace.

For analog circuits the evolution of technology is not as beneficial. Thus, there is a trend to move signal processing functions from the analog domain to the digital one, which, besides allowing for a higher level of accuracy, provides savings in power consumption and silicon area, increases robustness, speeds up the design process, brings flexibility and programmability, and increases the possibilities for design reuse. In many applications the input and output signals of the system are inherently analog, preventing all-digital realizations; at the very least a conversion between analog and digital is needed at the interfaces. Typically, moving the analog-digital boundary closer to the outside world increases the bit rate across it.

In telecommunications systems the trend to boost bit rates is based on employing wider bandwidths and a higher signal-to-noise ratio. At the same time radio architectures in many applications are evolving toward software-defined radio, one of the main characteristics of which is the shifting of the analog-digital boundary closer to the antenna.

Because of these trends, there is an urgent need for data converters with increasing conversion rates and resolution. A part of this needed performance upgrade comes

with the evolution technology, but often the demand is higher than this alone can provide. Thus, there is still room, and a need, for innovations in circuit design.

The increasing integration level leads to systems with a smaller number of chips, the ultimate goal being a single chip solution, the system on a chip (SoC). This means that analog and digital circuits have to live on the same silicon die, which brings additional challenges in analog design, such as mixed signal issues and limitations in the choice of technology. Data converters are inherently mixed signal circuits and face the same challenges on a smaller scale even without going as far as SoC. Furthermore, the evolution of technology has been driven by the microprocessor industry and hence does not always go in the best direction for the analog. However, the recent rapid growth of the wireless telecommunications devices market has given a boost to the development of advanced mixed signal technologies, such as silicon germanium-based BiCMOS.

The main challenges in data converter design are decreasing supply voltage, short channel effects in MOS devices, mixed signal issues, the development of design and simulation tools, and testability. In analog-to-digital converters (ADCs), they need to be met at the same time as the requirements for sampling linearity, conversion rate, resolution, and power consumption are becoming tighter.

This work concentrates on low voltage issues in ADCs by searching for and developing techniques and circuit structures suitable for today's and the future's low voltage technologies. In parallel, the increasing demands for ADCs have been answered by developing high-frequency high-linearity sampling techniques and applying them to ADC prototypes.

1.2   Organization of the Thesis and Research Contributions

The new ideas and circuits presented in this thesis have been partially reported in related publications [1]–[17]. The organization of the thesis is as follows:

Chapter 2 analyzes the effects of supply voltage and technology line width scaling on the signal-to-noise ratio, speed, and power consumption of switched capacitor circuits. Rather than searching for the fundamental limits, which has already been done by several authors, the analysis takes into account the limitations of real circuit structures and the properties of short channel MOS transistors.

Chapter 3 reviews the sample-and-hold operation, which is followed by a review of A/D conversion in Chapter 4. The chapter also introduces the most common high

speed Nyquist rate ADC architectures the main focus being on the pipelined ADC. Its optimal stage resolution is studied and methods of calibration discussed.

In Chapter 5 sample-and-hold circuit architectures are reviewed and compared.

Chapter 6 investigates the use of a MOS transistor as a sampling switch, focusing on high-frequency linearity issues. The known switch bootstrap circuits are reviewed and some improvements aimed at enhancing linearity and reducing signal feedthrough are proposed. One new circuit is also introduced.

Chapter 7 discuses opamp requirements in switched capacitor circuits. The most important single-stage and two-stage topologies are compared. How to meet the requirements of a high-speed, high-resolution pipelined ADC is demonstrated with a two-stage BiCMOS topology which has been developed [1].

Chapter 8 deals with clock generation issues in pipelined ADCs. Clock jitter is studied and an analysis of jitter in an inverter buffer is presented. The use of DLL for generating non-overlapping clock signals is also discussed.

In Chapter 9 an analysis of nonidealities in the double-sampling technique is carried out. A structure for eliminating one of them, the timing skew, is proposed [9].

Chapter 10 is a review of the switched opamp technique. Three new switchable opamps are proposed [8, 6, 2] and a passive input interface structure is developed to overcome the speed limitations of the existing active solutions [2].

In Chapter 11 another recently-introduced low voltage SC technique, based on unity gain reset opamps, is briefly described. The chapter also considers some low voltage current mirror, current source, and bandgap reference circuits [14].

Chapter 12 presents seven prototype circuits. Four of them have been designed and tested by the author himself, while the other three are results of the team work, in which the author has been the responsible project manager.

The first two are double-sampled S/H circuits [11, 10, 9], the latter of which uses the developed skew-insensitive sampling structure. The third prototype [4] is a 10-bit time-interleaved ADC, which is the result of work carried out jointly with Lauri Sumanen. The design of the S/H front-end, digital offset calibration, clock generation, and reference voltage buffering have been on author's responsibility and the architecture design as well as the development of the new comparator structure have been collaborative achievements. The fourth circuit [1] is an IF-sampling self-calibrated pipelined ADC, the design of which was performed by a team consisting of the author, Lauri Sumanen, and Tuomas Korhonen. Again the author's part of the work has been the design of the S/H front-end (including the opamp and bootstrapped switches) and clock generation. In the calibration the idea of adding an extra capacitor to the MDAC and the idea of doubling the gain of some back-end stages during calibration

are from the author. In the fifth prototype [16] a current steering DAC is combined with a current memory-based deglitcher. The implementation was carried out mainly by Jussi Pirkkalaniemi, the key ideas being from the author. The last two prototypes are pipelined ADCs realized with the switched-opamp technique. The first of them [8] is the first-ever reported application of this technique to a pipelined ADC. Its ideas and circuit structures are further developed in the second prototype [2].

Chapter 2

Low Voltage Issues

The whole history of integrated circuits has followed a trend of descending supply voltage. For a long time the de facto standard was 5 volts. The migration to a 3.3-volt supply in the mid-'90s started a trend in which almost every new process generation has a lower nominal supply voltage than its predecessor. Today the 0.25-μm generation uses a 2.5-V supply and, according to Semiconductor Industry Association's roadmap [18], it will be scaled down to 1.2 V by 2004 and to 0.9 V by 2008.

Table 2.1 shows process parameters for different technology generations. The data, including the effective channel length, supply voltage, oxide thickness, threshold voltage, and threshold voltage matching parameter, is collected from real processes. The table is reprinted from [19].

There are two main drivers for voltage scaling: technology and power. The shrinking technology feature size leads to lower break down voltages and thus supply voltage scaling is mandatory. Due to the ever-increasing integration level, which aims

Table 2.1 Technology data collected from different processes [19].

L_{min}	V_{DD}	T_{ox}	V_{th}	A_{Vth}
1.0	5.0	250	0.95	20
0.8	5.0	200	0.85	13
0.5	3.3	135	0.73	11
0.35	3.3	100	0.59	9.0
0.25	2.5	60	0.52	6.0
0.18	1.8	50	0.42	4.2
0.12	1.2	42	0.32	3.8
0.10	1.2	36	0.31	3.2
0.07	0.9	30	0.30	2.5

toward a system-on-a-chip (SoC), the power dissipation of a single chip tends to rise, which leads to severe heat problems and increased cooling system costs. On the other hand, the rapidly-growing market for portable battery-operated devices, such as laptops, PDAs, and cellular phones, demands high signal processing capacity together with low power dissipation.

The power dissipation of a logic gate is given by

$$P_{DIG} = \frac{1}{2} \cdot V_{DD}^2 \cdot C_L \cdot f_{clk} \cdot \alpha, \tag{2.1}$$

where V_{DD} is the supply voltage, C_L the load capacitance, f_{clk} the clock frequency, and α the switching probability. It is obvious that as a result of the quadratic dependence, the most effective way to reduce power consumption is to lower the supply voltage. Although it affects the circuit speed, every new technology generation comes with enhanced device characteristics and the possibility of increasing parallelism in the logic, which together more than compensate for the speed loss.

The situation regarding analog circuits is much more complicated and, as will be shown later, not so bright. The fundamental limits of power consumption in different types of analog circuits are discussed in several papers: switched capacitor filters in [20], continuous time filters in [21, 22], and data converters in [23]. The common finding in these papers is that there is a certain energy needed to present the signal and the resulting power is proportional to the signal or clock frequency and the desired signal-to-noise ratio, but not dependent on the semiconductor devices used or the supply voltage. In reality that fundamental limit cannot be reached or even approached because of the limitations of the technology (speed, noise, parasitic capacitances, etc.) and the circuit topologies. Low-voltage circuit techniques for different applications have been compared and analyzed, e.g. for filters in [24], for analog and digital video signal processing [25], and for various applications from sensor readout circuits to RF circuits in [26].

The remainder of this chapter concentrates on analyzing how supply voltage scaling affects the analog circuits, when the limitations of a CMOS technology are taken into account. The analysis concentrates on SC circuits and their most important building blocks: opamps and switches. The opamp is assumed to dominate power consumption, while the speed is determined by both the opamp and the switches.

To simplify the calculations, the transistor current is assumed to follow the square-law model, which does not describe very accurately the behavior of the deep-submicron transistors. The short channel effects and their impact on the results are, however, also discussed briefly.

2.1   Signal-to-Noise Ratio

One fundamental difference in analog signal processing compared to the digital is the significance of thermal noise [27], which sets a limit for the smallest distinguishable signal in the analog circuits. On the other hand, the supply voltage limits signal amplitude on the high side. The difference between these two boundaries determines the dynamic range, which is a key parameter in most systems. For a sinusoidal signal the peak signal-to-noise ratio is determined by

$$\sqrt{DR} = \sqrt{SNR_{max}} = \frac{V_{max}}{2\sqrt{2} \cdot \overline{V_n}} = \frac{V_{DD} - V_{margin}}{2\sqrt{2} \cdot \overline{V_n}}, \qquad (2.2)$$

where V_{max} is the maximum peak-to-peak signal amplitude and $\overline{V_n}$ the rms noise voltage. In real circuits the signal can never go all the way from the negative supply rail to the positive one; thus the V_{margin} in the second form of the equation. The required margin is highly dependent on circuit topology and somewhat dependent on current levels and the process parameters. It is typically some hundreds of millivolts at its minimum, ranging up to several volts at its maximum.

Lowering the supply voltage leads to decreased signal-to-noise ratio unless the noise level is scaled down simultaneously. What the cost is of keeping the DR constant in terms of circuit speed and current consumption is discussed next.

The thermal noise in CMOS circuits originates mainly from two sources: the resistors and the transistors. Which one dominates is circuit-dependent. In switched capacitor circuits the dominant noise source is typically the switch on-resistance and the rms noise voltage is given by

$$\overline{V_n} = k_1 \sqrt{\frac{kT}{C}}, \qquad (2.3)$$

where k is Boltzmann's constant, T the absolute temperature, C the sampling capacitor, and k_1 a constant dependent on circuit topology. For a given circuit the only way a designer can reduce the noise is to increase the capacitance.

The equation for gate-referred MOS transistor noise has the following form:

$$\overline{V_n} = \sqrt{\frac{4\gamma kTB}{g_m}}, \qquad (2.4)$$

where B is the noise bandwidth, g_m the transistor transconductance, and γ the noise excess factor. It will be shown later that in SC circuits g_m must be scaled linearly with the capacitance C to keep the circuit speed unaffected. This leads to the same

type of supply voltage dependency in circuit speed and power consumption as with the equation (2.3) and thus this case will not be discussed separately.

One hypothetical way to reduce thermal noise is cooling. However, substantial noise reduction is not obtained without extensive cooling. For example, cooling a circuit from room temperature (300 K) to the temperature of liquid nitrogen (77 K) reduces the noise by 6 dB or, alternatively, allows capacitor sizes to be reduced by a factor of four.

2.2   Circuit Speed

The speed of analog circuits is not usually directly dependent on the supply voltage. However, if the dynamic range is kept constant while decreasing the supply voltage, the capacitances have to be larger and thus the circuit speed is reduced. In addition, low-voltage circuit topologies are, in many cases, inherently slower than their high-voltage counterparts. Furthermore, the value of the parasitic drain and source junction capacitance increase as the substrate doping level increases and the reverse bias voltage decreases.

In switched capacitor circuits the maximum clock frequency is inversely proportional to the settling time, which is determined by slew rate and opamp bandwidth. For a single stage opamp the gain bandwidth product is given by

$$GBW = \frac{g_m}{2\pi C_L} = \frac{g_m}{k_3 C}, \qquad (2.5)$$

where g_m is the transconductance of the input transistor and C_L the load capacitance, which can be approximated to be proportional to the sampling capacitor C with circuit-dependent proportionality factor k_3. Solving C from (2.3) and (2.2), and substituting it to (2.5) yields

$$GBW = \frac{g_m(V_{DD} - V_{margin})^2}{k_3 DR \cdot k_1^2 kT}. \qquad (2.6)$$

Thus, when the settling time is dictated by the opamp GBW, the speed of an SC circuit decreases with the square of the supply voltage if DR is kept constant. It should be noted that the bandwidth loss can be compensated for by increasing the transconductance g_m.

Especially in moderate resolution circuits, the settling time can be dictated by the

slew rate, which is given by

$$SR = k_4 \frac{V_{max}}{T_S} = \frac{I_{SR}}{C_L}, \qquad (2.7)$$

where T_S is the clock period and I_{SR} the available slewing current. Solving T_S and substituting C_L and V_{max} from (2.3) and (2.2) yields

$$T_S = \frac{k_4 V_{max} C_L}{I_{SR}} = \frac{k_1^2 k_3 k_4 D R^2 kT}{I_{SR}(V_{DD} - V_{margin})}, \qquad (2.8)$$

which indicates that the attainable clock rate decreases linearly with the supply voltage unless the slewing current is increased.

2.3   Power Consumption

In the previous section it was shown that scaling down the supply voltage causes speed loss in the SC circuits unless the g_m of the opamp input transistor and/or the slewing current I_{SR} is/are increased. What the power consumption will be, if both the circuit speed and the *DR* are preserved while the supply voltage is reduced, is analyzed next.

2.3.1   Saturated MOSFET in Strong Inversion

Let us first look at the case where the circuit speed is limited by the opamp *GBW*. The opamp is a single stage operational transconductance amplifier (OTA) and its input transistor is realized with a MOSFET biased in the saturation region. Thus, the transconductance is

$$g_m = \sqrt{\frac{I_D \mu C_{ox} W}{L}} = \sqrt{\frac{I_D \mu C_G}{L^2}}, \qquad (2.9)$$

where I_D is the drain current, μ the carrier mobility, C_{ox} the gate oxide capacitance, and W and L the channel width and length. The second form of the equation is written using the expression for the gate capacitance $C_G = C_{ox}WL$. The gate capacitance appears between the opamp input and the ground and thus has an effect on the transfer function. It can be shown (see Appendix B) that in SC amplifiers (and integrators) there exists an optimum gate capacitance $C_{G,opt}$, which minimizes the settling time. This optimum is proportional to the sampling capacitor: $C_{G,opt} = k_5 C$, where k_5 is a circuit-dependent proportionality factor.

Now equation (2.5) for *GBW* can be rewritten

$$GBW = \frac{\sqrt{\mu k_5}\sqrt{I_D}}{k_3 L \sqrt{C}}. \tag{2.10}$$

Solving I_D yields

$$I_D = \frac{k_3^2 L^2 GBW^2 C}{\mu k_5}. \tag{2.11}$$

Using this the power consumption can be calculated to be

$$P_{SC,si} \propto I_D V_{DD} = \frac{k_3^2 k_1^2 L^2 GBW^2 DR \cdot kT}{\mu k_5} \cdot \frac{V_{DD}}{(V_{DD} - V_{margin})^2}, \tag{2.12}$$

where the second form is obtained by substituting C from (2.3) and (2.2). Next it is necessary to ask, if V_{margin} depends on V_{DD}. An assumption valid in most cases is that V_{margin} is determined by the saturation voltage $V_{dsat} = V_{GS} - V_T$ of the amplifier output stage transistors. An analysis carried out in Appendix C shows that, if the technology is fixed, V_{dsat} will be constant.

So it can be seen that for a given technology (fixed L) the power tends to increase when the supply voltage is decreased.

2.3.2 Saturated MOSFET in Weak Inversion

In many low-power and low-to-medium speed circuits the transistors are biased in the weak inversion region, where the transconductance is linearly dependent on the drain current and independent of the transistor aspect ratio. The power consumption can be calculated as before, resulting in

$$P_{SC,wi} \propto \frac{k_3 k_1^2 GBW \cdot DR \cdot k^2 T^2 n}{q} \cdot \frac{V_{DD}}{(V_{DD} - V_{margin})^2}, \tag{2.13}$$

where n is the subthreshold slope factor and q the charge of an electron. The parameter n is only slightly dependent on the bias point and the technology line width, being approximately 1.3 in a bulk silicon technology and 1.0 in a fully depleted silicon on insulator (SOI) technology [28].

2.3.3 Slew Rate Limited Power Consumption

When the SC circuit speed is slew rate limited, the power consumption is proportional to the slewing current and can be calculated using (2.8):

$$P_{SC,sr} \propto \frac{k_3 k_4 k_1^2 DR \cdot kT \cdot V_{DD}}{T_S (V_{DD} - V_{margin})}. \tag{2.14}$$

It is interesting to note that V_{margin} is the only term in the equation that has some technology dependency.

2.3.4 Technology Impact

How do the previous equations change when the voltage is scaled along with the technology? At least down to the 0.07-μm generation, the maximum supply voltage scales roughly linearly with the technology line width (see Table 2.1). Taking this into account by replacing the L proportionality with V_{DD} proportionality, the equation (2.12) for strong inversion can be rewritten

$$P_{SC,scaledL} \propto \frac{V_{DD}^3}{(V_{DD} - V_{margin})^2}. \tag{2.15}$$

The analysis in Appendix C yields that V_{margin} is proportional to V_{DD}^m, m ranging from 1 to 1.5. Substituting this into the previous equation results in a curve that decreases linearly with V_{DD} for $m = 1$ and decreases even faster if m is larger.

The technology scaling does not change the equations for weak inversion and slew-rate determined power except through V_{margin}. When $m = 1$ the slew-rate limited power is constant.

What makes power reduction possible in strong inversion is the fact that supply voltage scaling forces one to increase sampling capacitor size, resulting, as a positive side effect, in a larger optimum transistor gate capacitance. At the same time, the scaled-down technology offers more transconductance for a fixed current and fixed gate capacitance. In practice, the transconductance up-grade is obtained by increasing the W/L ratio, which moves the bias point toward weak inversion. Eventually, the transistor enters into weak inversion, where transconductance is not dependent on the aspect ratio and thus the power starts to increase according to (2.13).

2.3.5 Power Consumption: Summary and Conclusions

Figure 2.1 illustrates how the power consumption of thermal noise-limited SC circuits depends on the supply voltage, according to the previous equations. It is clear that, for a given technology, reducing the supply voltage increases the power consumption, the limiting factor being either opamp bandwidth or slew rate. When a scaled technology is utilized, with its nominal supply voltage, the slew rate-limited power is independent
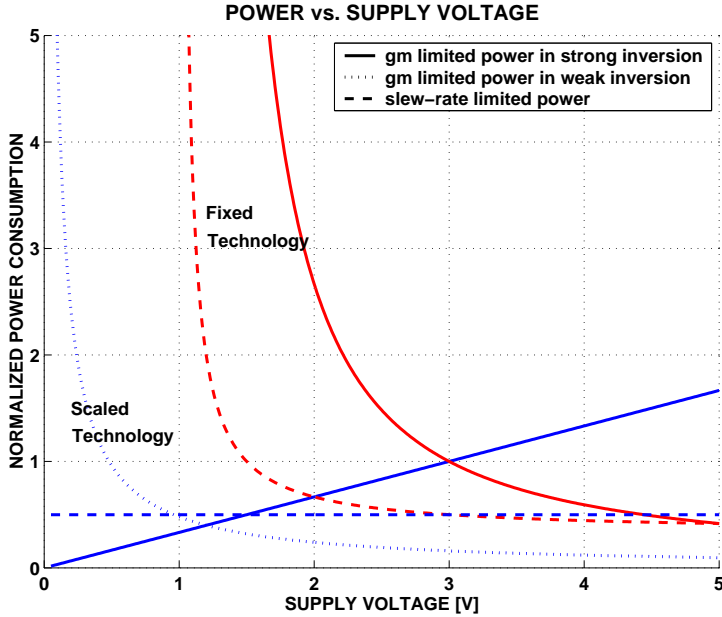
Figure 2.1 The effect of supply voltage scaling on the power consumption an SC circuit. A case where the technology line width is scaled along the supply voltage is compared to a case where the technology is fixed.

of the supply voltage, and the opamp bandwidth-limited power consumption decreases with the supply when the transistors are in strong inversion and increases when they are biased in the weak inversion. In practice, the power follows the curve which is largest and, as a result, there is an optimum technology after which further scaling leads to increased power consumption. The analysis presented in [22] shows a similar trend in continuous time circuits.

The analysis has not taken account of the short channel effects, the most important of which is velocity saturation. The carrier velocity saturates when the $V_{GS} - V_T$ voltage reaches a certain value, which is several volts for long channel devices but decreases with technology scaling. In velocity saturation the transistor current does not follow the square law; instead it is given by [29]

$$I_D = v_{sat} C_{ox} W \left(V_{GS} - V_T \right). \tag{2.16}$$

Consequently, the transconductance is independent of the current. As the technology is scaled down, the strong inversion region—between the weak inversion and the velocity saturation—gets narrower and eventually vanishes altogether [30, 31].

In Appendix B it is shown that for a given current and fixed L the minimum settling time of an SC amplifier is obtained in the strong inversion region and, when it is absent, at the point where the weak inversion region turns into the velocity saturation region. In the latter case the power consumption follows the weak inversion curve.

2.4   Matching

In most low-to-medium resolution circuits capacitor sizes are not limited by thermal noise but the matching. For example, in a 10-bit pipelined ADC with 2-V signal range the thermal noise sets the capacitor size below 0.1 pF, which is considerably smaller than typical capacitor sizes in 10-bit ADCs. In the signal transfer function capacitor ratios multiply voltages. Thus, the errors resulting from the mismatch do not depend on the absolute voltage values, making possible the scaling down of the supply and the signal range without changing the capacitor size. This results in a linear power reduction in opamp-bandwidth-limited power consumption and an even larger reduction in the slew rate-limited power consumption, even without technology scaling. At some point the thermal noise floor will be reached and capacitor size has to be increased.

Another type of mismatch error is the comparator and opamp offset voltage, originating from transistor threshold voltage and β ($= C_{ox}\mu W/L$) mismatch. In the signal transfer function the offset appears in an additive manner, thus causing larger relative error if the signal range is decreased. A widely accepted model for the threshold voltage mismatch of two transistors is given by [32, 33]

$$\sigma^2(\Delta V_T) = \frac{A_{Vth}^2}{WL} + S_{Vth}^2 D^2 \tag{2.17}$$

and for the β mismatch by

$$\sigma^2\left(\frac{\Delta\beta}{\beta}\right) = \frac{A_\beta^2}{WL} + S_\beta^2 D^2. \tag{2.18}$$

In most analog circuits the last term, which depends on the distance D between the devices, is small compared to the first one in both the equations and thus can be neglected.

In today's technologies the threshold voltage mismatch is the dominant source of offset in typical analog circuits. From Table 2.1 it can be seen that the parameter A_{Vth} scales linearly with the line width, and thus technology scaling improves the matching.

The β mismatch, however, does not improve significantly with the technology, and consequently at some point it starts to dominate [34], putting an end to the trend of improving matching. Although V_T matching improves with technology, the decreasing supply voltage and the increasing oxide capacitance result in worse matching with a given circuit speed and power consumption [35, 34].

Offset voltages are a major concern in flash and folding-and-interpolating ADCs, while pipelined ADCs are fairly robust against offsets. It has to be remembered that since mismatch is time invariant, it is fundamentally different from thermal noise. Therefore, various techniques can be used to reduce, correct, and calibrate the errors originating from mismatch. These techniques can be analog, digital, or mixed signal.

2.5 Operational Amplifiers

When opamps are employed as building blocks in SC circuits, they can almost always be used in the inverting feedback configuration, where the signal swing in the opamp input is very small. Thus, the input structure does not limit the signal range. The opamp output stage, in contrast, sees the full signal swing and ultimately sets the maximum limit for it.

Single-stage opamp topologies, such as the telescopic opamp or the folded cascode opamp, are regarded as the fastest and most power-efficient structures for integrated applications. In achieving the required high DC gain they rely on cascoding, which limits the output signal swing. When two or more gain stages are used, the required DC gain can be obtained without cascoded structures in the amplifier output. This type of opamp, for example, the traditional Miller opamp, requires somewhat more power to achieve the same bandwidth as single-stage opamps.

In Figure 2.2 two output stages and corresponding signal swings are shown: a cascoded output stage on the left and a rail-to-rail output stage on the right. Despite its name, the latter circuit cannot provide true rail-to-rail signal swing, since there is one transistor between the signal and the supply rail. This is, however, as close as it is possible to go: thus, the name is justified. For proper operation the transistors have to be in saturation, which requires the drain-source voltage to be at least equal to the saturation voltage V_{dsat}, which depends on the current level and technology line width. In typical opamp designs the values are some hundreds of millivolts. In practice, in addition to the V_{dsat}, some extra voltage margin has to be reserved to achieve robustness against inaccurate biasing and to get a decent output impedance.

The reduction in signal swing in the cascode stage is twice that in the rail-to-rail stage. It is obvious that as the supply voltage gets lower the margin eats an increasing
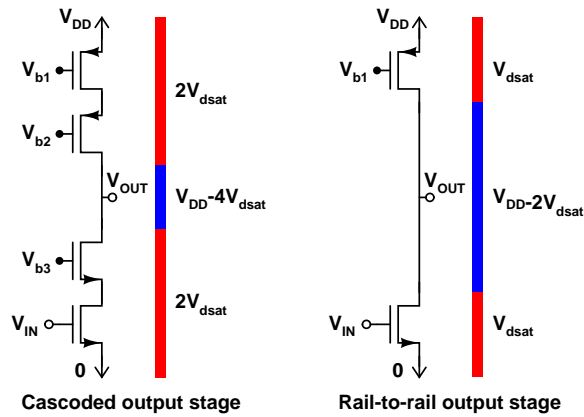
Figure 2.2 Output signal swing in different output stages.
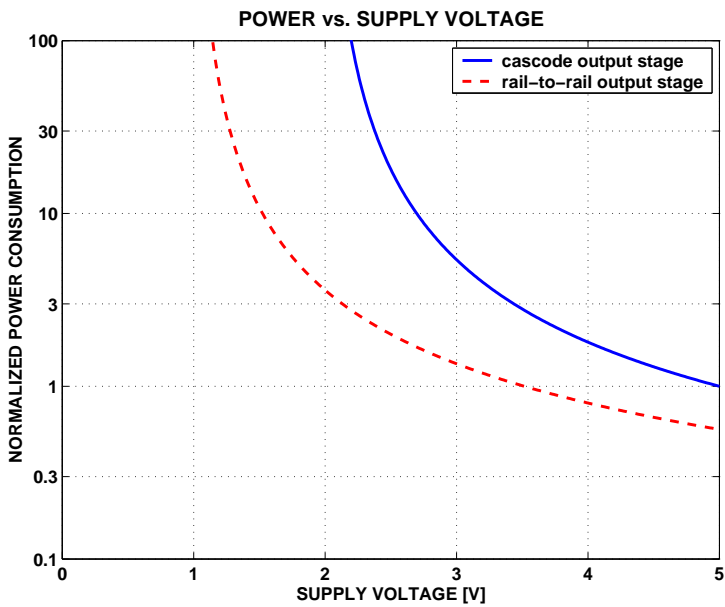
Figure 2.3 Power consumption of circuits using different types of opamp.

portion of the signal range. How this affects power consumption is illustrated in Figure 2.3. There the power consumption, given by equation (2.12), is plotted against the supply voltage for both output stages. The power is normalized in such a way that for the cascode circuit it is 1 at 5 V. The voltage margin in the rail-to-rail circuit is set to 1.0 V and in the cascode circuit to 2.0 V.

At 5 V the circuit with the cascode stage dissipates 78% more power, which will probably be compensated for in a more efficient opamp topology. When the supply voltage is lower, the penalty grows bigger, being 400% at 3 V, which undoubtedly favors the opamp with the rail-to-rail output stage. Thus, opamp topology has a significant effect on power consumption. Consequently, multi-stage opamps with rail-to-rail output stage are preferable choices in low-power low-voltage designs.

2.6   MOS Switches

Another crucial building block in the SC circuits is the switch. An ideal switch has infinite resistance when it is open and zero resistance when it is closed. At high supply voltages (5 V and higher), a MOS transistor has been a good enough approximation of that. On the other hand, the finite on-resistance of a closed MOS switch has already caused problems in some applications with a 3-V supply and the problems are believed to get worse as the supply is scaled down. Whether these problems are due to increased circuit performance requirements or to the technology scaling is investigated next.

The on-resistance of a MOS switch can be written as

$$R_{ON} \approx \frac{L}{W \mu C_{ox}(V_{GS} - V_T)},$$ (2.19)

where V_{GS} is the transistor gate-source voltage and V_T the threshold voltage. The equation is valid when $V_{GS} \geq V_T$; with smaller gate-source voltages the resistance is infinite. To turn the switch properly on, its gate-source voltage has to be V_T plus some overdrive to make the on-resistance small enough.

A single-transistor switch cannot conduct over the whole rail-to-rail signal range, since, for example, an nMOS switch, whose gate is tied to V_{DD}, cuts off when the signal level is raised within a threshold voltage of V_{DD}. This is illustrated in Figure 2.4 (a), where the inverse of the on-resistance is plotted against the signal level. The whole range can be covered by putting an nMOS and a pMOS transistor in parallel to form a CMOS switch or a transmission gate (Figure 2.4 (b)). The on-resistance has its largest value in the mid-range between the supplies, when the overdrive voltage
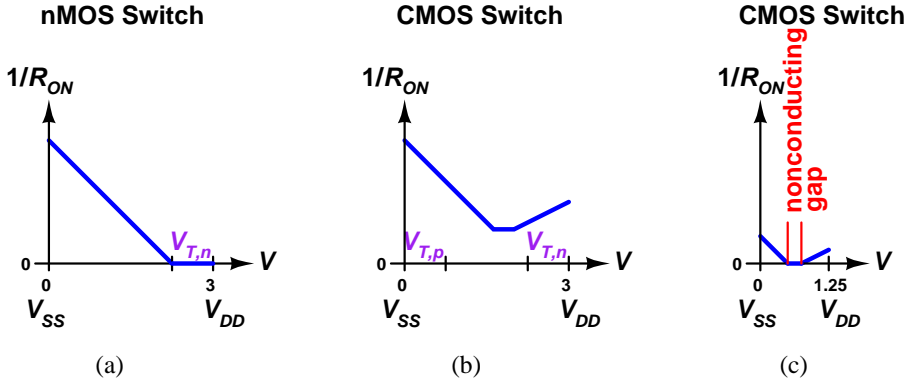
Figure 2.4 Inverse of the switch on-resistance as a function of the signal voltage for an nMOS (a) and a CMOS (b) switch with high supply voltage and a CMOS switch with low supply voltage (c).

is approximately $V_{DD}/2 - V_T$. Thus, the maximum on-resistance, which is given by

$$R_{ON,max} \propto \frac{L}{C_{ox}\mu W\left(V_{DD} - 2V_T\right)}, \tag{2.20}$$

has a strong correlation with the supply voltage.

When V_{DD} becomes smaller than $V_{T,n} + V_{T,p}$ a non-conducting gap appears in the mid-supply range (Figure 2.4 (c)). In most applications the switch becomes useless much earlier as a result of too-large on-resistance, which makes the settling times long. Another problem, especially in S/H circuits, is the signal-dependent nature of the resistance, which causes harmonic distortion when sampling continuous time signals. Consequently, the resistance has to be much smaller than in the case of a constant on-resistance.

What is the effect of technology scaling on the on-resistance? Because of increasing leakage current in digital circuits the threshold voltage cannot be scaled down linearly with the supply voltage, but the scaling rather follows a square root function. The threshold voltage values from 0.35-μm generation to 0.07-μm generation in Table 2.1 fit quite accurately to $0.32 \cdot \sqrt{V_{DD}}$. Using this and assuming a linear dependence between the line width and the supply voltage equation (2.20) can be rewritten

$$R_{ON,max} \propto \frac{V_{DD}^2}{C_G\left(V_{DD} - 0.64 \cdot \sqrt{V_{DD}}\right)}, \tag{2.21}$$

where $C_G = C_{ox}WL$.

Remembering that reducing the signal range increases the capacitances, the correct

parameter to look at, is not the resistance, but the time constant, which is given by

$$\tau = RC \propto \frac{C}{C_G} \cdot \frac{V_{DD}^2}{\left(V_{DD} - 0.64 \cdot \sqrt{V_{DD}}\right)}. \tag{2.22}$$

The parasitic capacitances of the switch are proportional to C_G and it is reasonable to assume that they can be allowed to grow at the same rate as the capacitance C. Thus, the first term in the equation is constant and the second one decreases linearly with the supply at high V_{DD} values, but starts to rise rapidly when the supply goes below 0.9 V.

Methods to reduce the on-resistance can be divided into two categories: technology-based and circuit-based. The technology-based methods reduce the resistance by lowering the transistor threshold voltage, while the circuit-based methods increase the overdrive voltage.

One method is the use of a dual-V_T process, which provides two types of transistors, with either a high or low threshold voltage. The idea of this technology is to improve the speed of digital logic by using low-V_T transistors in critical places but at the same time keeping the leakage current small with high-V_T transistors. In analog designs the low-V_T transistors can be employed as switches. There are, however, at least two problems with this technology: it is not a mainstream technology—at least today—and so it costs more and is not necessarily available with analog extensions (capacitors and resistors). Circuits with low threshold switches may also suffer from charge leakage.

Another technology that may alleviate the switch problem is the silicon-on-insulator (SOI) technology. SOI MOSFETs have inherently lower leakage current than their bulk CMOS counterparts and so with the same channel length the threshold voltage of an SOI transistor is typically smaller. The future will show whether this technology will become mainstream.

The methods that improve the switch by means of circuit techniques increase the overdrive voltage from $0.5V_{DD} - V_T$. The switch transistor gate-source voltage can always be raised to at least V_{DD} without violating the technology specifications. A switch with a gate-source voltage of V_{DD} will have an overdrive voltage of $V_{DD} - V_T$, which is a substantial improvement on the earlier situation. One technique that can be used to realize this is gate-voltage bootstrapping. In this the voltage at the switch transistor gate is capacitively boosted above the supply voltage so that it follows the signal voltage with an offset equal to V_{DD}.

Another completely different circuit technique that also results in an overdrive of $V_{DD} - V_T$ is the switched opamp technique [36]. In this the SC circuit and the opamp are modified in such a way that all the switches can operate against ground or virtual
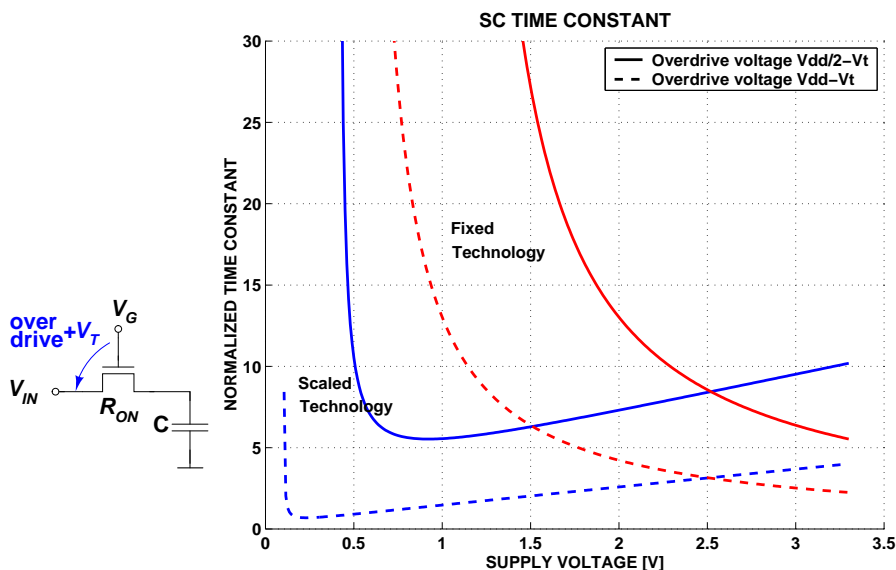
Figure 2.5 Time constant of a switched capacitor as a function of supply voltage.

ground. Consequently, the switch does not see any voltage swing and the overdrive is always the maximum.

In Figure 2.5 the time constant is plotted as a function of the supply voltage with two different overdrive voltages, with and without technology scaling. It can be seen that technology scaling yields smaller SC time constants, even without any special techniques, down to a supply of 0.9 V. Increasing the overdrive brings a significant advantage by lowering the on-resistance in all supply voltages and extending switch usability to lower voltages.

Both bootstrapping and the switched opamp technique are discussed and analyzed more deeply in the following chapters.

2.7   Conclusions

In today's world, where digital circuits offer more speed and capacity with reduced power consumption with every new technology generation, the expectations for analog circuits are similar. It has, however, been clear for a long time that the benefits of technology scaling are not so great for them, mainly because of the decreasing supply voltage. During recent times the question has rather been whether or not the technology scaling *degrades* the performance of the analog circuits.

The analysis of SC circuits presented in this chapter shows that there is, indeed,

some benefit to be gained from technology scaling, at least for the next few technology generations. This, however, is probably not enough for the increased requirements and expectations that exist. Thus, there is an urgent need for techniques at the architectural and circuit levels for alleviating the problems associated with low voltage and more effectively taking advantage of the technology.

At the circuit level it is of the utmost importance to maximize the signal swing, which has a large impact on opamps and switches. In noise-limited circuits, using a supply voltage smaller than the maximum allowed does not bring any advantage at the circuit level. When accuracy is limited by capacitor matching, a lower supply may be justified.

Technology scaling is best exploited by doing things digitally. This does not only mean moving the signal processing functions from the analog domain to the digital one but also combining both techniques in realizing the system blocks. A good example is that of digitally self-calibrated pipelined ADCs, where technology scaling has made possible the incorporation of more and more complex digital calibration algorithms into ADCs. Similar types of techniques, to be used against mismatch and nonlinearity, can probably be developed for other applications as well in order to improve performance and permit more robust analog structures.

Chapter 3

Sample-and-Hold Operation

3.1 S/H Basics and Performance Metrics

The main function of a sample-and-hold (S/H) circuit is to take samples of its input signal and hold these samples in its output for some period of time. Typically, the samples are taken at uniform time intervals; thus, the sampling rate (or clock rate) of the circuit can be determined.

The operation of an S/H circuit can be divided into sample mode (sometimes also referred as acquisition mode) and hold mode, whose durations need not be equal. In hold mode, the output of the circuit is equal to the previously sampled input value. In sample mode, the output can either track the input, in which case the circuit is often called a track-and-hold (T/H) circuit, or it can be reset to some fixed value. In some circuits the output is held over the whole period of the sampling clock. This is achieved by having separate circuitry to perform the sampling and the holding operations.

The most common terms and performance metrics used in conjunction with S/H circuits ([37, 38]) are briefly introduced in the remainder of this section. Which of

Track-and-Hold   **Sample-and-Hold**
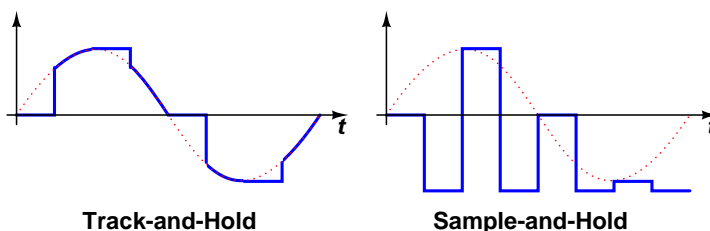
Figure 3.1 Output waveforms of different S/H circuits.
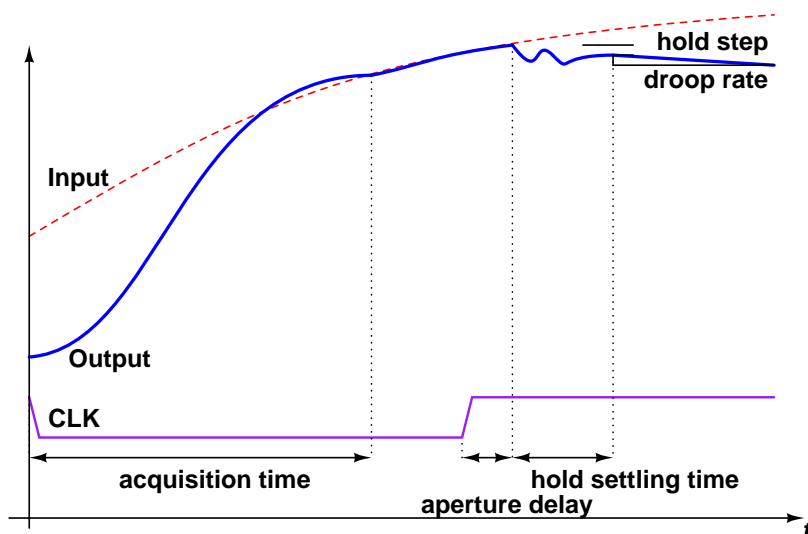
Figure 3.2 Time domain S/H circuit performance metrics.

them are more important than the others greatly depends on the application of the S/H circuit. The technology utilized also has some effect on which parameters are usually given in circuit specifications. To fully characterize an S/H circuit, specifications both in the time domain and in the frequency domain have to be defined. Unfortunately, some of the terms used are not well-established and thus the definitions in different sources may be contradictory.

The *acquisition time* is the time from the command to switch from hold mode to sample mode to the moment when the circuit is ready to take a new sample, i.e. it tracks the input. Acquisition time is one of the parameters that define the maximum achievable sampling rate.

The *aperture time* or *aperture delay* is the fixed time from the sampling command to the moment when the sample is actually taken.

Random variation in the sampling moment is known as *aperture uncertainty* or *aperture jitter*.

The *hold mode settling time* determines the time from the sampling moment to the moment when the circuit output has settled within the specified accuracy of its steady state value. If the S/H circuit is used in front of an ADC, the ADC can digitize the S/H circuit output at that moment. The hold mode settling time has a major impact on the maximum sampling rate of the S/H circuit.

The signal may leak from the circuit when in hold mode. The rate of change in output that results from this is specified by the *droop rate*.

The *hold step* or *pedestal error* is usually defined for track-and-hold circuits. It is the difference in the output value at the end of the tracking and during hold mode. The pedestal may be signal-dependent and thus produce harmonic distortion.

During hold mode the signal at the circuit input may couple to the output. The fraction of the input signal seen at the output is specified by the *hold mode feedthrough*.

Usually, S/H circuits have a unity gain (i.e. the amplitude of the output signal is equal to the amplitude of the input signal), but other gain values can be used as well. The *gain error* determines the deviation of the gain from the nominal value.

The *dynamic range* is the difference in decibels between the maximum allowed input voltage and the minimum input voltage that can be sampled with a specified level of accuracy.

Nonlinearity in the S/H circuit causes distortion. Measured with a sinusoidal input signal, the *total harmonic distortion* (*THD*) is the ratio of the sum of error energy in the frequencies harmonically related to the input frequency to the signal energy at the fundamental frequency. The THD can be given as a percentage or in decibels. In sampled data systems, aliasing complicates the identification of the harmonic frequencies in the spectrum.

The *spurious free dynamic range* (*SFDR*) is the ratio of the largest spurious frequency and the fundamental frequency.

The *signal-to-noise ratio* (*SNR*) is the ratio of noise energy to signal energy.

The *signal-to-noise-and-distortion ratio* (*SNDR* or *SINAD*) is the ratio of all error energy to signal energy. Quite often the term signal-to-noise ratio is used although SNDR is actually meant.

When an S/H circuit is employed in the ADC front-end it is meaningful to speak of *resolution*, which is expressed as a number of bits. Resolution is just another way to express the SNDR for the maximum input signal and it is obtained by $(SNDR - 1.76)/6.02$.

3.2   Spectra of Sampled Signals

An ideal S/H circuit takes samples of an input signal at uniform intervals T. In the time domain this corresponds to multiplying the signal by an impulse train

$$y(t) = x(t) \cdot \sum_{n=-\infty}^{\infty} \delta(t - nT), \tag{3.1}$$
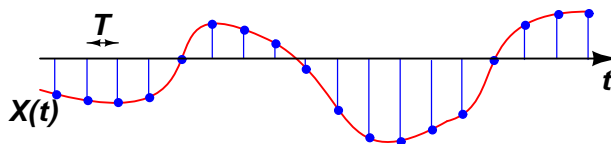
Figure 3.3 Sampling in time domain.

where $\delta(t)$ represents Dirac's delta function. The result is a train of impulses whose values correspond to the instantaneous values of the input signal.

The spectrum of the sampled signal is a convolution of the input spectrum and the spectrum of the impulse train, which is also an impulse train

$$Y(f) = X(f) \star \sum_{n=-\infty}^{\infty} \frac{1}{T}\delta(f - \frac{n}{T}).$$ (3.2)

This is illustrated in Figure 3.4, where f_S is the sampling frequency and B the signal bandwidth. The resulting spectrum is the original spectrum plus an infinite number of images of the original spectrum centered at multiples of the sampling frequency. The figure also clearly shows that as long as the bandwidth of the input signal is less than half of the sampling frequency the images do not overlap and thus the original signal can be restored by filtering. If this condition—known as the Nyquist criterion—is not satisfied, a part of the image spectrum is aliased into the desired signal band, causing irreversible distortion. Because of this, the input signal usually has to be band limited before sampling in order to avoid the aliasing of noise and other unwanted signals present outside the desired signal band. In sub-sampling (or under-sampling) the aliasing is utilized to sample high frequency narrow-band signals. There, a signal band around some multiple of the sampling frequency is aliased to the baseband, which actually corresponds to down conversion. This can be used in radio receivers to digitize the intermediate frequency (IF) signal, using a relative narrow band ADC. In principle, the signal can be sub-sampled even at radio frequency (RF), but noise aliasing and sampling clock jitter limit performance and prevent the use of technique in most systems [39].

3.2.1 Spectrum of a Sampled and Held Signal

In practice, the output waveform of a sampling circuit cannot be a train of infinitely narrow impulses. In most practical implementations the sample is held in the output of the circuit until the next sample is taken (Figure 3.5). In that case the circuit is known as a sample-and-hold (S/H) circuit. Sometimes the output tracks the input for
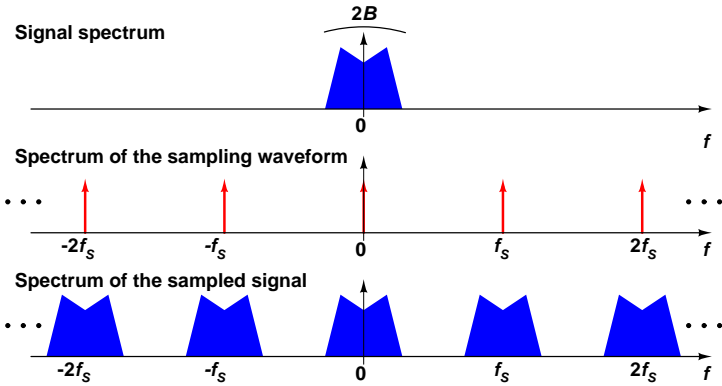
Figure 3.4 Spectrum of a sampled signal.
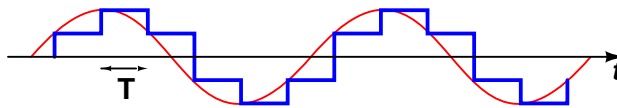
Figure 3.5 A Sampled-and-held signal.

half of the sample period and is held in the sampled value for the other half. This type of circuit can be called a track-and-hold (T/H) circuit. However, inconsistent terminology is quite often seen.

If the signal processing after a S/H circuit is performed in discrete time, which is, for instance, the case in ADCs with a front-end S/H circuit, the spectrum is the ideal periodic spectrum as in Figure 3.4. On the other hand, if the output waveform of the S/H circuit is used as a continuous time signal the spectrum is different. A deglitcher preceding a DAC is a typical example of this kind of circuit. A similar situation can also occur when measuring S/H circuits.

The time domain representation of a sampled-and-held signal is a convolution of the sampled signal (3.1) and a square pulse

$$y(t) = \left[x(t) \cdot \sum_{n=-\infty}^{\infty} \delta(t - nT) \right] \star \Pi \left(\frac{t}{T} - \frac{1}{2} \right), \tag{3.3}$$

where $\Pi(t/T - 1/2)$ denotes a square pulse from $t = 0$ to $t = T$. In the frequency domain the convolution corresponds to multiplication and thus the spectrum of the sampled-and-held signal is the spectrum of a sampled signal multiplied by the spectrum of the square pulse, which has the form of $\sin(x)/x$. Using this well-known rela-
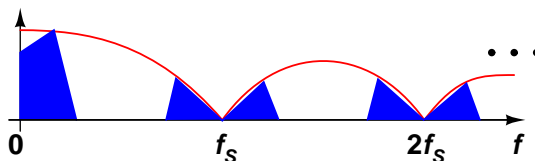
Figure 3.6 Spectrum of sampled-and-held signal.

tionship the spectrum of a sampled-and-held signal can be written as

$$Y(f) = e^{-j\pi fT} \cdot \frac{\sin(\pi fT)}{\pi fT} \sum_{n=-\infty}^{\infty} X(f - \frac{n}{T}). \tag{3.4}$$

A power spectrum of this form is shown in Figure 3.6. In many cases the sinc attenuation is not tolerable and the signal has to be either predistorted before the hold operation or corrected after it.

3.2.2 Sampling Function

When ideal impulses (Dirac's delta functions) are used to describe operations in analog continuous time signal processing, one should be on the alert. It turns out that it is impossible to realize a circuit performing the sampling according to (3.1). In practice, a circuit cannot pick the instantaneous value of its input signal, but rather it takes a weighted average of the input during a time window around the sampling moment. Mathematically, this is equal to integrating the product of the input signal and the sampling function from minus infinity to plus infinity in the time domain. For a single sample this can be written as follows:

$$y(t_0) = \int_{-\infty}^{\infty} x(t)h(t - t_0)dt, \tag{3.5}$$

where t_0 is the sampling instant and $h(t)$ the sampling function. The same for an infinite sequence of samples is

$$y(nT) \quad = \quad \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} x(t)h(t - nT)dt \tag{3.6}$$

$$= \quad \sum_{n=-\infty}^{\infty} x(nT) \star h(-nT) \tag{3.7}$$

$$= \quad [x(t) \star h(-t)] \sum_{n=-\infty}^{\infty} \delta(t - nT). \tag{3.8}$$

The integral in (3.6) can be identified as a convolution resulting (3.7). This can be interpreted as a sampled form of the convolution integral. In (3.8) the same is presented using the Dirac's delta function. Utilizing this, the frequency domain signal can be easily obtained with the Fourier transform:

$$Y(f) = [X(f) \cdot H(-f)] \star \sum_{n=-\infty}^{\infty} \delta(f - \frac{n}{T}). \tag{3.9}$$

This shows that in the frequency domain the effect of the sampling function is seen as a multiplication by the conjugate of the Fourier transform of the sampling function. Since the Fourier transform of an impulse is 1 the equation (3.9) is consistent with (3.2). A more realistic sampling function than Dirac's delta function is a triangular pulse, whose Fourier transform is $\sin^2(\pi f T_b/2)/(\pi f T_b/2)^2$, where T_b is the width of the base of the triangle. In general it can be said that a real sampling function always adds a low-pass filtering effect to the sampling operation. The modeling of the limited tracking bandwidth can also be included in the sampling function [40].

3.3 Noise Issues in S/H Circuits

3.3.1 kT/C Noise

Any sampling circuit can be considered as consisting of at least a switch and a capacitor. The switch always has some finite on-resistance which generates thermal noise. The power spectral density of this noise is the well-known $4kTR$ V^2/Hz, where k is Boltzmann's constant, T the absolute temperature, and R the resistance. The noise in the voltage sample is the resistor noise filtered by the low-pass circuit formed by the sampling capacitor and the switch on-resistance. Integrating the resistor noise spectral density weighted by the low-pass transfer function yields the mean square noise voltage on the capacitor

$$\sigma^2 = 4kTR \int_0^\infty \frac{1}{1 + (f2\pi RC)^2} df \tag{3.10}$$

$$= \frac{4kTR}{(2\pi RC)^2} \Big|_0^\infty 2\pi RC \arctan(f2\pi RC) = \frac{kT}{C}. \tag{3.11}$$

By looking at the result it becomes obvious why this noise is often referred to as kT/C noise. An interesting point is that the noise voltage does not depend on the value of the switch on-resistance, and thus the only parameter which can be used to control the noise is the value of the sampling capacitor. Although the desired signal bandwidth

is typically at least an order of magnitude smaller than the noise bandwidth of the sampling circuit, the sampled noise is still determined by (3.11). This is due to the fact that the sampling operation aliases all the noise energy into the Nyquist band.

In ADCs a common requirement is that thermal noise power is smaller than the power of the quantization noise, which can be shown to be $LSB^2/12$. This sets the lowest limit for capacitor value C as follows

$$C > \frac{kT \cdot 12}{LSB^2} = \frac{kT \cdot 12}{2^{-2N}V_{FS}^2}, \tag{3.12}$$

where N is the number of bits and V_{FS} the voltage corresponding to the ADC full scale. Sometimes the requirement is more stringent, allowing only 1 dB of SNR degradation, which changes the factor 12 in the equation to 46.3. According to (3.12), in the case of 1-volt full-scale voltage, the capacitor values required for 10- and 16-bit resolution are 0.052 pF and 210 pF respectively, which indicates that the capacitor values for 16-bit resolution begin to be too large for practical integration. To overcome this, a popular solution in high-resolution applications is to use an oversampling ADC architecture, in which the capacitor size can be reduced linearly with the oversampling ratio.

3.3.2   Jitter in Sampling Clock

Random variation of the sampling instant is known as jitter. It originates from clock generator phase noise and sampling circuit noise. How the jitter is transformed to the amplitude error in the sampled voltages can be understood as follows: the error in the sampled voltage is equal to the change in the input voltage between the ideal sampling instant and the actual sampling instant. The voltage change in turn is proportional to the jitter and the rate of change of the input signal, i.e. its derivative. For a sinusoidal input the derivative is the cosine function multiplied by the corner frequency, which means that the voltage error is proportional to the frequency and the amplitude of the input signal. It can be shown [41] that the signal-to-noise ratio limited by jitter can be written as

$$SNR = -20\log(2\pi f \Delta t), \tag{3.13}$$

where f is the frequency of the input signal and Δt the rms value of the jitter. It can be seen that increasing the amplitude of the input signal does not improve SNR, since it also increases voltage error. Jitter is studied further in Chapter 8.

3.3.3   Other Noise Sources

Most S/H circuits need a buffer amplifier or an opamp, at least when in hold mode. The internal noise sources of the amplifier add in power to the thermal noise of the switch on-resistances. In passive sampling the noise is band limited by the RC time constant of the sampling circuit. When an amplifier contributes to the circuit transfer function, which is usually the case in hold mode and, in some closed-loop S/H architectures, also in sampling mode, its finite bandwidth is likely to be the dominant band-limiting factor. To reduce the amount of aliased noise the bandwidth of the amplifier should be kept as small as is permitted by the settling requirements [42]. This is important, since if the S/H circuit is followed by an ADC the S/H circuit noise during hold mode is also aliased due to the sampling performed by the ADC.

In addition to white noise, S/H circuits also suffer from flicker or $1/f$ noise. However, in high-speed applications (a clock frequency of several megahertz), the white noise typically dominates. This is further enhanced by noise aliasing. If the $1/f$ noise becomes a problem there are several techniques, such as correlated double sampling (CDS) or chopper stabilization to get rid of it [42].

3.4   Basic S/H Circuit Architectures

In hold mode an S/H circuit remembers the value of the input signal at the sampling moment, and thus it can be considered as an analog memory cell. The basic circuit elements that can be employed as a memory are capacitors and inductors, of which the capacitors store the signal as a voltage (or charge) and the inductors as a current. In addition to the inductor, a current memory needs a switch that is a good short circuit when it is closed. Similarly, a switch which is a good open circuit in its off-state is needed for a voltage memory. Since capacitors and switches with a high off-resistance are far easier to implement in a practical integrated circuit (IC) technology than inductors and switches with a very small on-resistance, all sample-and-hold circuits are based on voltage sampling. There also exists current mode S/H circuits, but they always include voltage-to-current and current-to-voltage converters which allow the sampled quantity to be voltage.

S/H circuit architectures can roughly be divided into open-loop and closed-loop architectures. The main difference between them is that in closed-loop architectures the capacitor, on which the voltage is sampled, is enclosed in a feedback loop, at least in hold mode.
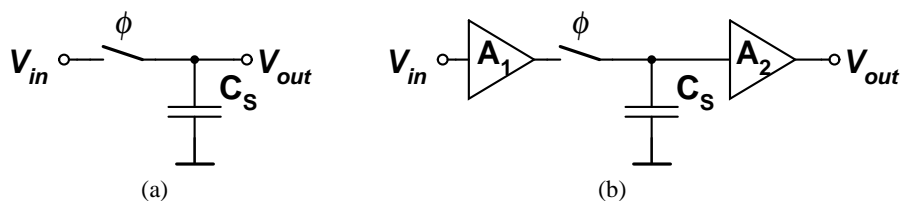
Figure 3.7 A simple S/H circuit (a) and a practical S/H circuit (b).

3.4.1 Open-Loop Architectures

The simplest S/H circuit consists of a switch and a capacitor (Figure 3.7 (a)). In sample mode the switch is closed and the voltage on the capacitor tracks the input signal. During the transition to hold mode the switch is opened and the input voltage value at the switch opening moment stays on the capacitor. This circuit, however, is impractical since it is not capable of driving any load. Therefore a buffer has to be used to drive the load. An input buffer may also be needed to adjust the signal level to one suitable for the switch and to reduce hold mode feedthrough. An S/H circuit with an input and an output buffer is shown in Figure 3.7 (b).

The main advantage of this open-loop S/H architecture is its high speed. Accuracy, however, is limited by the harmonic distortion arising from the nonlinear gain of the buffer amplifiers and the signal-dependent charge injection from the switch. These problems are especially emphasized with a MOS technology.

3.4.2 Closed-Loop Architectures

A well-known technique to improve linearity is the utilization of negative feedback. The feedback can be used internally in the buffer amplifiers in an open-loop architecture like the one in Figure 3.7 (b). However, this does not help with switch-induced distortion. The signal-dependent charge injection can be avoided by operating the switch at a constant potential, which can be realized by enclosing the switch in a feedback loop to create a virtual ground. Figure 3.8 shows a basic closed-loop S/H circuit following this idea [43, 44].

As a result of feedback the output tracks the input in sample mode. The switch is connected to the virtual ground provided by the second operational amplifier and thus it introduces only a constant error charge. When the switch is opened the global feedback loop is broken and the input voltage is sampled into the capacitor C_H. The capacitor is permanently connected in a feedback loop around the second operational amplifier, which is used as a buffer both in track mode and hold mode.
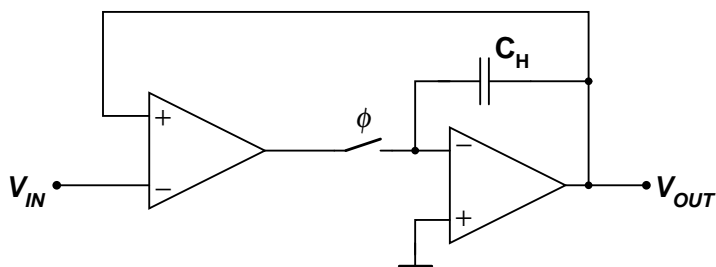
Figure 3.8 A basic closed-loop S/H circuit.
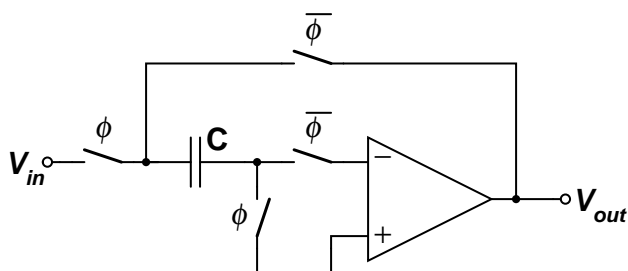
Figure 3.9 A switched capacitor S/H circuit.

Since the feedback loop encloses two opamps in tracking mode the circuit has to be heavily compensated in order to avoid instability. This naturally reduces the speed of the circuit. Another potential disadvantage is hold mode feedthrough via the parasitic input capacitances of the first operational amplifier.

A closed-loop S/H architecture, commonly used in switched capacitor (SC) circuits and referred to as flip-around S/H, is shown in Figure 3.9. It performs the sampling passively, i.e. it is done without the opamp, which makes signal acquisition fast. In hold mode the sampling capacitor is disconnected from the input and put in a feedback loop around the opamp, as in the circuit shown in Figure 3.8. Signal-dependent charge injection from the switches is avoided by a technique called bottom plate sampling, which relies on special timing of the switch control signals. This technique is discussed in more detail in Section 6.3.

Chapter 4

A/D Converters

4.1 A/D Conversion

Analog-to-digital (A/D) conversion can be separated into two distinct operations: sampling and quantization. Sampling transforms a continuous time signal into a corresponding discrete time signal, while quantization converts continuous amplitude distribution into a set of discrete levels, which can be expressed with digital code words. Figure 4.1 shows the principle of A/D conversion.

Some A/D converter (ADC) architectures, the flash for instance, can perform sampling and quantization simultaneously, and in some ADCs, which are targeted for DC signals, no sampling is needed at all. In high performance ADCs, however, sampling and quantization are usually separated to make it possible to optimize the circuitry for both tasks without compromises. Furthermore, the performance of many ADC architectures, which do not necessarily need a separate sampling circuit, can often be improved by adding one.
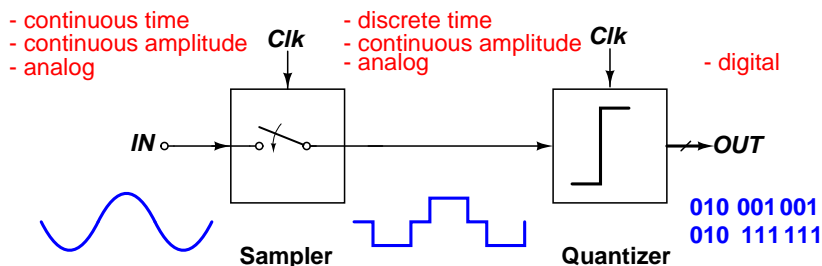
Figure 4.1 Principle of A/D conversion.
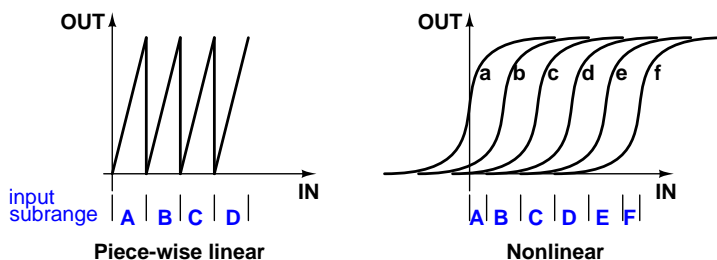
Figure 4.2 Piece-wise linear and nonlinear analog preamplification.

The sampling operation has already been discussed in Chapter 3 and the architectures of S/H circuits will be investigated in Chapter 5. In the remainder of this chapter the quantization operation is studied in more detail and the most common high-speed ADC architectures are introduced. There is also a bias is toward those architectures most suitable for CMOS technologies. Furthermore, the focus is on high-speed and medium- to high-resolution (8 bits or more) ADCs. Oversampling ADCs are not discussed.

4.1.1 Direct Quantization

The most straightforward way to perform quantization is to compare the signal to a reference with a comparator. One comparison yields a one-bit result, telling whether the signal is larger or smaller than the reference. Thus, to get greater accuracy i.e. a larger number of bits, more comparisons are needed. For a single signal value, several successive comparisons can be made against different reference levels (successive approximation architecture), yielding at most as many bits as there are comparisons. Alternatively, with several parallel comparators the signal can be compared against many references at once (flash architecture), which gives a multibit result in one comparison phase. In the case of two-step flash these two methods are combined, requiring fewer comparators than flash ADC and being faster than successive approximation ADC.

In these architectures quantization accuracy is, to a large extent, determined by comparator accuracy. The larger their number and speed requirement are, the more difficult their realization at a reasonable cost in area and power becomes. As a result, architectures using analog signal preprocessing prior to quantization have been developed.

4.1.2 Quantization After Analog Preprocessing

The preprocessor provides gain, which relaxes the comparator accuracy requirement. Since the signal range in ADCs is typically comparable to the supply voltage, a linear amplifier cannot provide significant gain. Thus, the preprocessor either has a piece-wise linear transfer function (algorithmic and pipeline ADCs) or produces several shifted nonlinearly amplified signals, out of which one is selected automatically (flash with distributed preamplification) or depending on a coarse approximation of the signal (folding-and-interpolating ADC). The two different types of transfer functions are illustrated in Figure 4.2. The preprocessing can be realized in a continuous or discrete time domain.

Preprocessing also helps in reducing the number of comparators. This can be seen, for instance, from the piece-wise linear transfer function, which folds several incoming subranges into one output range, reducing the number of comparators in the final quantization by a factor equal to the number of folds. To prevent information loss, a coarse quantization has to be performed prior to or in parallel with the preprocessing to distinguish the correct input subrange. The accuracy requirement of the coarse quantization can be relaxed by using redundancy (overlapping subranges), and thus it does not affect the final quantization.

4.1.3 ADC Figures of Merit

Some of the parameters used with ADCs are also found in the case of S/H circuits and hence have already been presented in the previous chapter. A more comprehensive presentation of ADC specifications than the one that follows can be found, for example, in [37].

The *sampling rate* tells how many samples the ADC can process in a time unit and the *latency* how many clock cycles there are between the sampling instant and the moment when the digital code is available at the ADC output.

The accuracy of the conversion is specified by its *resolution*, which is given in bits. Often, resolution is a synonym for the *number of output bits*, not all of which necessarily carry any valuable information. Thus, the real accuracy can be specified by an *effective number of bits (ENOB)*, which is just the *signal-to-noise-and-distortion ratio (SNDR)* expressed in bits.

Ideally, the SNDR is limited by the finite precision of the quantization, which leaves an error between the original non-quantized signal and the quantized signal. Often, this error is treated statistically and referred to as quantization noise as a consequence of its uniform amplitude distribution and virtually flat spectral density. Quanti-
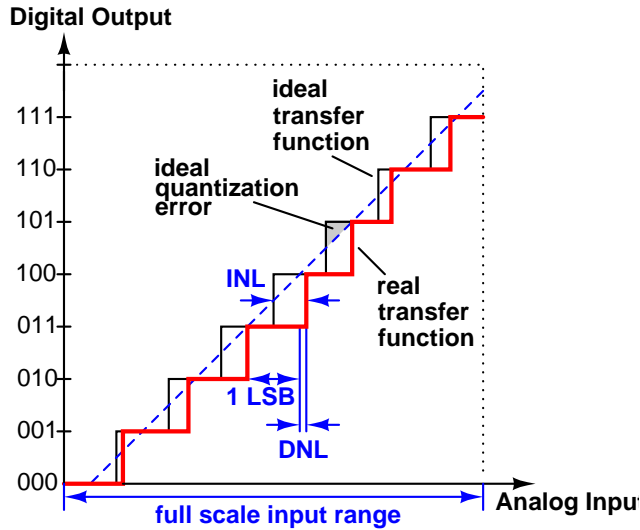
Figure 4.3 Errors in ADC transfer function.

zation noise limits the SNDR (with full-scale signal) to $6.02 \cdot N + 1.76$ dB, where N is the resolution. In a practical ADC thermal noise, disturbing signals, and transfer function errors make the SNDR smaller. From this measured SNDR value the effective number of bits can be calculated:

$$ENOB = \frac{(SNDR - 1.76)}{6.02}. \tag{4.1}$$

To gain more information about how the error energy is spectrally distributed, *spurious free dynamic range (SFDR)* and *total harmonic distortion (THD)* are often specified. Their definition is the same as with S/H circuits, and unless otherwise specified they are obtained with a full-scale input signal. The *effective resolution bandwidth (ERB)* is the input signal frequency, below which the ENOB is less than half a bit worse than its DC value.

The full-scale signal range is expressed in volts. From it, the voltage step corresponding to the least significant bit (LSB) can be calculated: $LSB = V_{FS}/2^N$. The static errors in the transfer function are specified with *differential non-linearity (DNL)* and *integral non-linearity (INL)*, both of which are referenced to the LSB. They can be either presented graphically as a function of the output code, or only the maximum value, which is a single number, can be given. The DNL is the error in the step size between two adjacent quantization levels, which is ideally 1 LSB. The INL is the cumulative DNL, and it is equal to the deviation from a straight line drawn between the
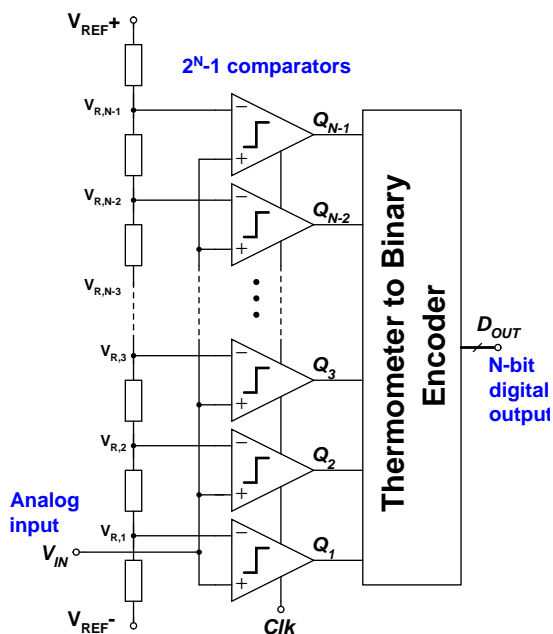
Figure 4.4 N-bit flash ADC.

end points of the transfer function. The DNL and INL are illustrated in Figure 4.3. In addition to these the ADC may also contain *offset* and *gain error*, which are not shown in the figure.

4.2 Flash ADC

Flash ADC, which is the fastest and one of the simplest ADC architectures, is shown in Figure 4.4. It performs $2^N - 1$ -level quantization with an equal number of comparators. The reference voltages for the comparators are generated using a resistor ladder, which is connected between the positive ($V_{REF}+$) and the negative ($V_{REF}-$) reference voltage determining the full-scale signal range. Together the comparator outputs form a $2^N - 1$ -bit code, where all the bits below the comparator whose reference is the first to exceed the signal value are ones, while the bits above are all zeros. This so-called thermometer code is converted to N-bit binary word with a logic circuit, which can also contain functions for removing bit errors (bubbles).

Since the input signal is directly connected to the inputs of the comparators, flash architecture is very fast; the speed is only limited by the comparators. Thus, the fastest reported ADCs are realized with this architecture. Flash ADC also has very

low latency—typically one to two clock cycles—which allows it to be utilized in applications using feedback (e.g. gain control loop).

The most prominent drawback of flash ADC is the fact that the number of comparators grows exponentially with the number of bits. Increasing the quantity of the comparators also increases the area of the circuit, as well as the power consumption. Thus, very high resolution flash ADCs are not practical; typical resolutions are seven bits or below.

Other issues limiting the resolution and speed include nonlinear input capacitance, location-dependent reference node time constants, incoherent timing of comparators laid out over a large area, and comparator offsets. To manage the offsets the utilization of auto-zeroing comparators is often necessary. Alternatively, the offsets and input capacitance can be reduced by means of distributed preamplification combined with averaging [45, 46], and possibly also with interpolation [47], which can both be considered as exponents of the signal preprocessing discussed in the previous section.

In averaging each comparator is preceded by a preamplifier, whose output is coupled to the outputs of the adjacent preamplifiers via a resistive averaging network. As a result, the input signal for a comparator is not produced by its own preamplifier alone, but it is a weighted average of the outputs of the preamplifiers in a small neighborhood. Comparator offset is reduced by the preamplifier gain and the preamplifier offset is an average of the random offsets of all the amplifiers participating in the amplification.

Not every comparator needs to have a preamplifier of its own; instead, some (typically every other or three out of four) amplifiers can be eliminated and the missing signals generated by means of interpolation. Neither averaging nor interpolation reduces the number of the comparators, and thus it does not significantly extend flash architecture toward higher resolutions.

Recently, the main application of flash ADCs has been in disk drive read channel circuits and local area network interfaces. Typically, six-bit resolution with a sampling rate of several hundred megahertz is required. Even gigahertz rates seem to be within the reach of state-of-the-art CMOS technologies [48, 49].

4.3   Subranging ADC

One way to reduce the number of comparators in flash ADC is to perform the quantization in two phases [50]. First a coarse quantization determines the subrange where the signal lies and then, in the second phase, the quantization is performed inside this range. Consequently, the number of comparators can be reduced from $2^N - 1$ to
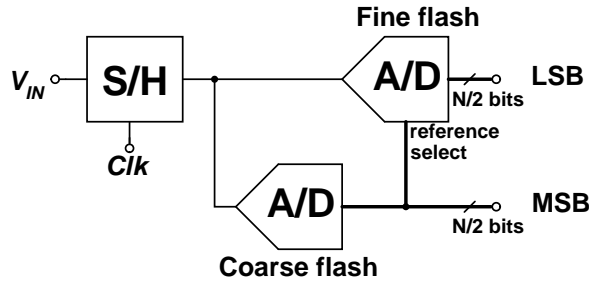
Figure 4.5 Subranging ADC.
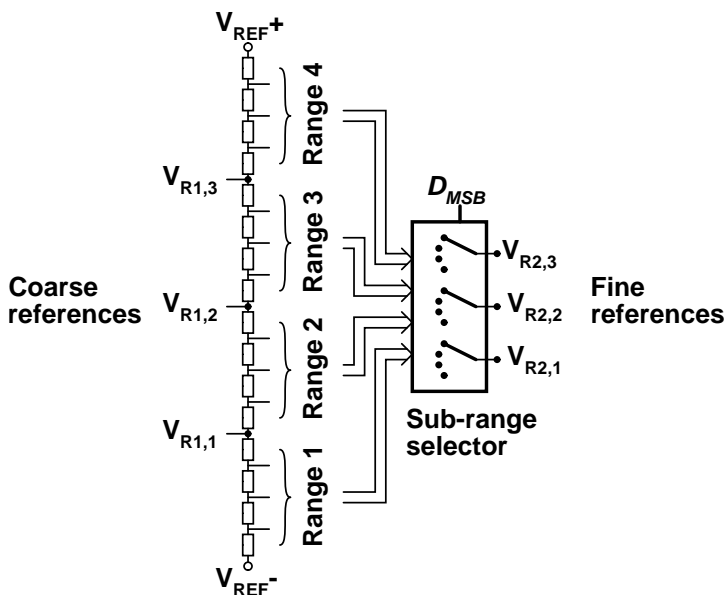
Figure 4.6 Subrange selection.

$2 \cdot \left(2^{N/2} - 1\right)$, assuming that the number of bits determined in each phase is the same. In an 8-bit ADC this means 30 comparators instead of 255. This type of architecture is known as either two-step flash or subranging architecture.

A block diagram of a subranging ADC is shown in Figure 4.5. There, the input signal is first sampled with a sample-and-hold circuit, which guarantees that both of the flashes have the same input signal. The first flash resolves the most significant bits (MSBs), which are also utilized for determining the coarse input range, according to which the reference voltages for the second flash are selected. A principal implementation of this selection, employing a common resistor ladder for both stages, is shown in Figure 4.6. For example, when the coarse A/D conversion tells us that the signal is
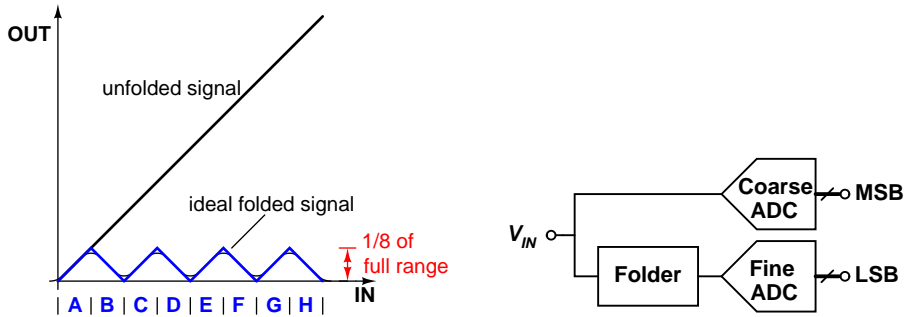
Figure 4.7 Folding principle and a block diagram of a folding ADC.

between the first stage reference voltages $V_{R1,2}$ and $V_{R1,3}$, the taps in range 3 are used as fine references.

Subranging architecture allows for higher resolutions than the flash method, thanks to the reduced number of comparators. It cannot achieve as high a speed because of the S/H circuit that is needed and the fact that the fine ADC can operate only after the coarse result is available. The architecture does not alleviate the comparator accuracy requirement, unless inter-stage gain is introduced. Then, however, it should rather be referred to as a two-stage pipeline. Adding redundant comparators, which are connected to the taps in the verges of the neighboring subranges, to the fine ADC, can be used to relax the comparator accuracy requirements in the coarse flash.

The applications of subranging architecture are most often found in video signal acquisition, where 10-bit resolution with a 20–40 MS/s sampling rate is typically needed. With CMOS technology a sampling rate as high as 100 MS/s with 8-bit resolution has been reported in [51], while 12-bit resolution at 54 MS/s has been achieved in [52].

4.4 Folding-and-Interpolating ADC

4.4.1 Folding

Figure 4.7 shows the idea of folding. In this the input signal range is divided into eight subranges (A to H), at the boundaries of which the signal is folded up or down. As a result the output signal range is squeezed into one-eighth, reducing the required number of comparators by the same amount. In addition, folding allows signal amplification back to the full scale (not shown in the figure), which helps in reducing the effect of comparator offsets and noise.
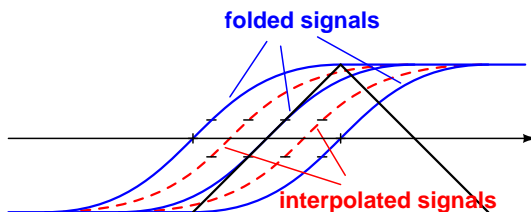
Figure 4.8 One linear signal replaced with several nonlinear ones.

The concept of folding is very similar to subranging, the major difference being that folding does not need a priori knowledge of the input subrange. As a result, no S/H circuit is needed, which leads to faster operation. Although the subrange need not be known at the time of folding, it is needed in the forming of the final conversion result. Thus, a coarse ADC is used in parallel with the folding circuit, as shown in the principal block diagram in Figure 4.7.

In practice, realizing a transfer function with the triangle wave shape is very difficult, since especially the sharp corners tend to become smoothed. This problem can be solved by producing several versions of the folded signal, each shifted a different amount in the x-direction, and using only the linear part of each curve. This is illustrated in Figure 4.8, where five nonlinear curves are used instead of one linear one. The little marks around the zero crossings show which portion of each curve is utilized. All the comparators responsible for detecting the signal in this range are connected to the circuit producing that particular curve. Often, the number of curves is increased up to the point where they equal the number of comparators. As a result, there is only one comparator per curve and it only has to detect the signal zero crossing, making the linearity of the curve unimportant.

4.4.2 Interpolation

The folder is a rather complex circuit, and thus the number of them should not be too large. Fortunately, it can be reduced by generating part of the signals by interpolating between two signals produced by adjacent folders. In Figure 4.8 two of the signals (dashed lines) are generated with interpolation. The interpolation happens in the y-direction, as a result of which the resulting curve is not a perfect substitute for a real folded signal. In fact, only one zero crossing can be interpolated between two curves without an error. In practice, however, a larger number of interpolated signals can be used, when the folded signals are sufficiently linear and the shifting of the zero crossings is taken into account in the interpolation network.
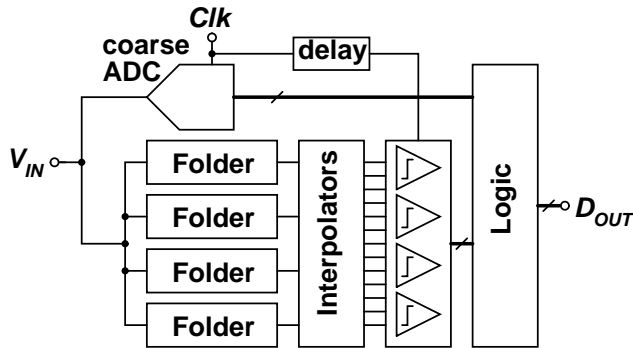
Figure 4.9 Folding-and-interpolating ADC.
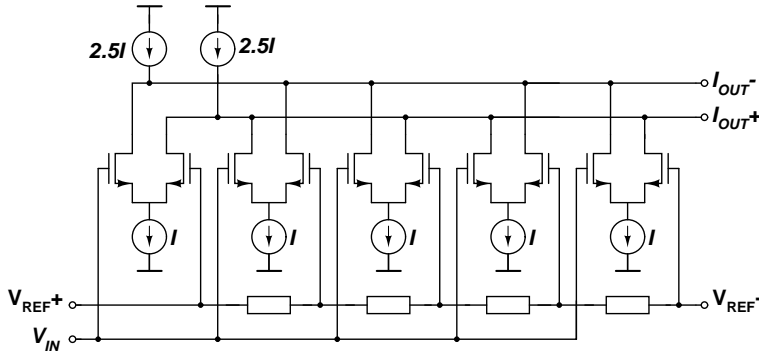
Figure 4.10 CMOS folding amplifier with differential current output.

4.4.3 ADC Architecture

A block diagram of folding-and-interpolating ADC is shown in Figure 4.9. In this four folders are followed by an interpolation network, to which the comparators are connected. A coarse flash ADC determines the correct input subrange, which is used with the comparator outputs to form the final output code in the logic block. Different internal signal propagation delays in the coarse ADC and in the folding and interpolating circuitry need to be taken into account to guarantee simultaneous timing in the two circuits. Adding redundant folded or interpolated signals increases tolerance to timing errors and comparator inaccuracy in the coarse ADC.

A CMOS implementation of the folding amplifier is shown in Figure 4.10 [53]. It produces a four-times-folded differential current signal with five differential pairs. The interpolation between two folders can be realized with a resistor ladder.

The folding-and-interpolating architecture can achieve high sampling rates. This is mostly thanks to open loop circuitry (folding amplifiers do not use feedback) and

continuous time operation. The circuit has a relatively small silicon footprint and it can typically be realized with a digital CMOS process, i.e. without precision capacitors.

4.4.4 Limitations and Improvements

A well-known problem in this architecture is the fact that because of the folding the frequencies of the internal signals are much higher than the frequency of the incoming signal. As a result the performance typically begins to degrade at relatively low signal frequencies. The problem can be alleviated by using an S/H circuit in front of the converter, which, however, often diminishes the speed advantage. In distributed track-and-hold [54] each differential pair in the folding amplifier has its own preamplifier with track-and-hold, which has less stringent specifications than a front-end S/H circuit would have. As a result, a higher speed can be achieved.

The folding-and-interpolating architecture was originally developed for bipolar technology, which is ideal for realizing accurate open-loop circuits thanks to good V_{BE} matching and high transconductance of the bipolar transistor. On the other hand, the offset voltages in MOS transistors are the main obstacle to increasing resolution. Thus, techniques such as averaging [46, 55] and self calibration [56] are used to reduce offset sensitivity.

The folding can be carried out in many cascaded stages, minimizing the number of folds per stage [46, 55]. Consequently, the number of differential pairs connected to the input is reduced, which allows for the biasing of the transistors to a larger gate-source voltage, which increases the speed via increased transconductance. The capacitive load is also reduced, giving an additional increase in circuit speed.

The throughput of the folding-and-interpolating architecture can be improved, at the cost of latency, by using pipelining, which can be realized by combining cascaded folding with the distributed T/H as demonstrated in [57] and [56]. Both of these designs also use subranging to reduce the number of folders.

The resolution of the folding-and-interpolating ADCs reported has typically been in the 8–10-bit region and the sampling rates from some dozens of megahertz to a hundred megahertz. As high as a 400-MS/s sampling rate has been achieved [58]— with 6-bit resolution, though. The resolution of CMOS realizations has been limited to 10 bits, with one exception [56], which uses background self-calibration to cancel the offsets of the folding amplifiers.

The folding amplifier structure based on differential pairs does not allow for very low-voltage operation, since it requires at least $V_T + 2V_{dsat}$ on top of the input signal swing. Consequently, many of the ADCs described in the references use a 5-volt
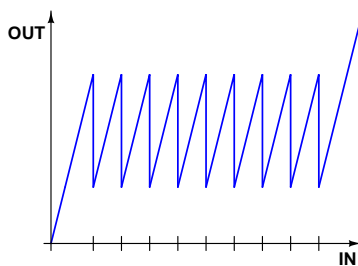
Figure 4.11 Residue transfer function of pipelined ADC.

supply and not a single one goes below three volts.

4.5 Pipelined ADC

Like folding-and-interpolating ADC, pipelined ADC also uses analog preprocessing to divide the input range into subintervals and to amplify the signal inside them, as seen from the transfer function shown in Figure 4.11. The realization of the preprocessing is, however, totally different. The architecture has evolved by making use of the strengths of the switched capacitor technique, which provides very accurate and linear analog amplification and summation operations in the discrete time domain. As a result the sawtooth-shaped transfer function can readily be realized. Because of the high quality of the preprocessing it is often used so extensively that the resolution of final A/D conversion is only one or two bits.

A predecessor of pipelined ADC is algorithmic (or cyclic) ADC, which performs the conversion (or processing) in several time steps. It uses a single processing unit, which works with the analog sample for a certain, say, m, number of clock cycles. In pipelined ADC, m of those units—usually called pipeline stages—are cascaded. Each stage processes the same sample only for one clock cycle, after which it passes it to the next stage for further processing. The end result and the latency are the same as in algorithmic ADC, but the throughput has increased m times, being now one conversion per clock cycle.

4.5.1 Pipelined A/D Conversion: Principle

The principle in pipelined (and algorithmic) A/D conversion is to find a set of reference voltages whose sum equals the signal sample being converted. This is realized by sequentially subtracting different reference voltages from the sample until the residue becomes zero, indicating that the sum of the subtracted references equals the original

sample value. An analogy can be found in weighing flour on a pair of scales using a set of weights. In ADC the residue is amplified between the subtraction steps in order to increase accuracy.

4.5.1.1   Operation of a Pipeline Stage

For each stage, the input signal is the output of the previous stage ($V_{OUT,i-1}$)—except for the first stage, for which it is the input voltage V_{IN}. The stage evaluates the incoming signal with a coarse A/D conversion, which yields the digital output code Q_i. Now, as the magnitude of the input signal is roughly known, the closest multiple of the reference voltage ($Q_i \cdot V_{R,i}$) can be subtracted from it and the resulted residue signal amplified by the stage gain G_i, yielding the output voltage

$$V_{OUT,i} = G_i \cdot V_{OUT,i-1} - Q_i \cdot V_{R,i}, \tag{4.2}$$

which is used as an input for the next stage. It should be noted here that the coarse A/D conversion needs only to be accurate enough to prevent residue voltage saturation after amplification. If this is satisfied, its accuracy does not affect the converter's final accuracy, unlike the precision of the amplification and the reference subtraction. Tolerance toward sub-ADC errors requires that the input range of the next stage and the set of its reference voltages have at least as much overhead as the largest allowed error is.

4.5.1.2   Forming the Output Code

In principle, the conversion can be continued for an infinite number of cycles, each of which increases the precision of the result. In reality, the result does not improve after some point as a consequence of component mismatch and noise. Once the desired number of conversion cycles has been gone through, the final conversion result is the sum of the subtracted reference voltages, referred to the input of the ADC:

$$V_{OUT} = \overline{Q_1} \cdot \overline{V_{R,1}} + \sum_{i=2}^{m} \left(\overline{Q_i} \cdot \overline{V_{R,i}} / \prod_{j=1}^{i-1} G_j \right), \tag{4.3}$$

where $\overline{Q_i}$ is the i:th stage digital output as a bit vector and $\overline{V_{R,i}}$ an equal length vector containing the stage's reference voltages. The digital code corresponding to (4.3) is obtained by replacing the reference voltages with their digital values. Typically, the final resolution is increased by performing an A/D conversion (without forming a residue) for the last residue signal. This is the only sub-A/D conversion whose
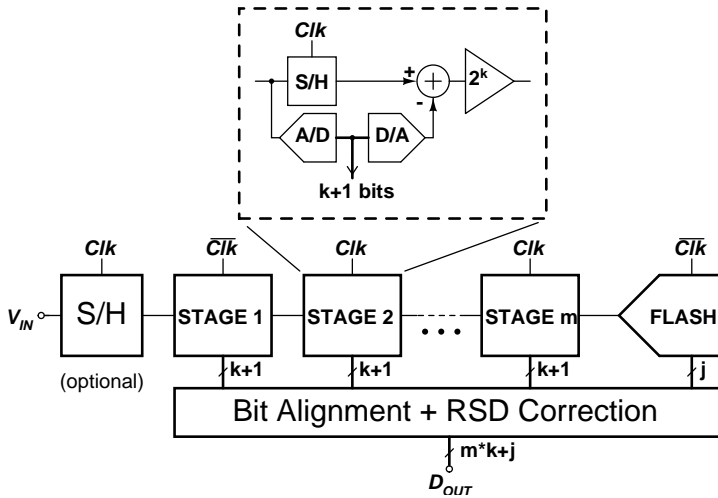
Figure 4.12 Pipelined ADC.

accuracy directly affects the accuracy of the conversion.

4.5.2 Pipeline Architecture

The block diagram of the pipelined ADC is shown in Figure 4.12. It consists of *m* stages, each of which produces *k* bits plus one redundant bit, which overlaps with the bits of the next stage. The last pipeline stage is followed by a flash ADC, providing *j* bits. As a result, the final resolution *N* is $m \cdot k + j$. In practice, the resolution of all the stages does not need to be equal and the redundant bit does not need to be a full bit (i.e. the number of possible output codes can be less than 2^{k+1}).

A functional block diagram of one stage is shown in the inset. The incoming voltage is sampled by the S/H circuit and simultaneously digitized by the sub-ADC. The result of the A/D conversion is immediately converted back to analog form and subtracted from the sampled-and-held signal. The resulting residue voltage is amplified by G_i, which is nominally equal to 2^k. In a switched capacitor realization the S/H operation, the D/A conversion, the subtraction, and the amplification are all performed by a single circuit block, called a multiplying analog-to-digital converter (MDAC), which consists of an opamp and a set of switched capacitors. The low resolution sub-ADC is usually a flash, consisting of a few comparators and logic gates.

A front-end S/H circuit is not necessarily needed, since the pipeline stage already contains an S/H circuit. When the input is a rapidly-changing signal, the relative timing of the first stage S/H circuit and the sub-ADC is critical and often relaxed with
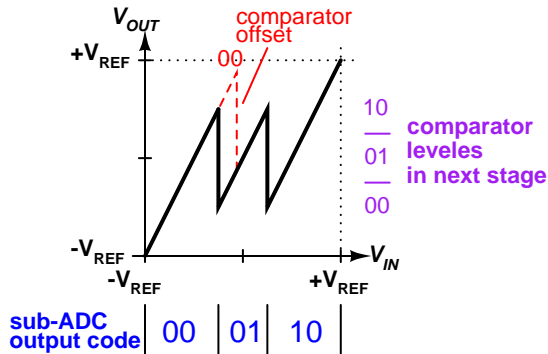
Figure 4.13 Transfer function of a 1.5-bit pipeline stage.

a front-end S/H circuit.

Consecutive stages operate in opposite clock phases and as a result one sample traverses two stages in one clock cycle. So, the latency in clock cycles is typically half the number of stages plus one, which is required for digital error correction. For feedback purposes, where low latency is essential, a coarse result can be taken after the first couple of stages. The different bits of a sample become ready at different times. Thus, digital delay lines are needed for aligning the bits.

The first monolithic pipelined ADC [59] was implemented with the switched capacitor technique, which has since become the standard in pipelined ADCs. Current mode approaches have also been tried [60, 61], but they have not been able to achieve the same performance.

4.5.3   RSD Correction

The simplest pipeline stage is a 1-bit stage with one redundant quantization level, as a consequence of which it is often referred to as a 1.5-bit stage. Its transfer function is shown in Figure 4.13. The signal range both in the input and in the output is from $-V_{REF}$ to $+V_{REF}$. Nominally, the comparator decision levels are set to $-V_{REF}/4$ and $+V_{REF}/4$ and the ADC output codes for the three regions are "00", "01" and "10".

Comparator offset can shift the decision level, as shown in the figure with the dashed line. Consequently, the ADC output code remains "00" instead of "01". The residue, however, stays in the input range of the next stage, which is all that matters, as discussed earlier. The final conversion result is the sum of all the subtracted reference voltages. Thus, if one stage subtracts a smaller reference than it nominally should, the subsequent stages compensate for this by subtracting larger references. How this

**Stage
output bits**

stage 1 b_{11} b_{10}

stage 2 b_{21} b_{20}

stage 3 b_{31} b_{30}

stage 4 b_{41} b_{40}

$+$

combined output d_4 d_3 d_2 d_1 d_0

$\Longleftrightarrow$

b_{11} b_{10} b_{31} b_{30} 0

$+$ 0 b_{21} b_{20} b_{41} b_{40}

d_4 d_3 d_2 d_1 d_0
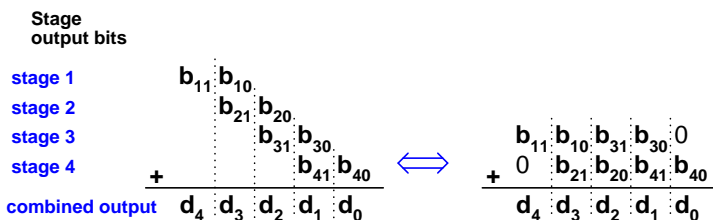
Figure 4.14 RSD correction in digital domain.

works in the case illustrated in the figure is shown next.

Due to quantization error the residue now lies in the "10" region (instead of the "00" or "01" region) of the next stage. Since the outputs of the stages are summed with one-bit overlap, the second most significant bit will be correct, regardless of the comparison error. When the codes of the following stages are taken into account, the third bit also gets corrected, and so on.

In the digital domain the correction is a simple addition, as illustrated in Figure 4.14. On the left is shown how the bits of the final result are obtained by summing the output bits of the stages with one-bit overlap. On the right the same bits are rearranged to show how the correction can be performed with a single adder. Since "11" codes are not possible (except in the last flash stage), the sum cannot overflow.

This widely-employed error correction method is referred to as redundant signed digit (RSD) correction, and was developed for algorithmic ADCs in [62] and [63] and utilized in pipelined ADC in [64]. Other related methods have also been used (e.g. [59]).

The redundancy allows for quantization errors as far as the residue stays in the input range of the next stage, which translates to $\pm V_{REF}/4$ maximum error in 1.5 bits/stage architecture. The errors can be static or dynamic; it is only essential that the bits going to the correction logic circuitry match those which are D/A converted and used in residue formation.

The same correction method can easily be expanded to larger resolution stages as well. As a minimum, one extra quantization level is required [65], but for maximum error tolerance the nominal number of comparators ($2^k - 1$) has to be doubled.

4.5.4 Switched Capacitor Realization

Switched capacitor implementation of an MDAC suitable for a 1.5-bit pipeline stage is shown in Figure 4.15. During clock phase ϕ it samples the input voltage into two nominally equally-sized capacitors C_1 and C_2. In the second phase the capacitor C_2 is
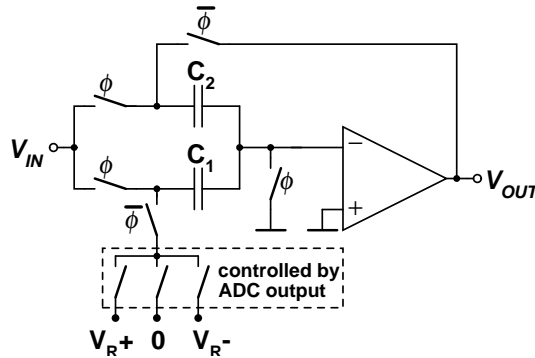
Figure 4.15 Switched capacitor MDAC for a 1.5-bit pipeline stage.

connected in a feedback loop around the amplifier and the capacitor C_1 to one of the three reference voltages, according to the output of the stage's sub-ADC. The resulting output voltage is given by:

$$V_{OUT} = \frac{C_1 + C_2}{C_2} \cdot V_{IN} + Q \cdot \frac{C_1}{C_2} \cdot V_{REF} = 2V_{IN} + Q \cdot V_{REF}, \qquad (4.4)$$

where Q is the ADC output code with possible values -1, 0, and $+1$. The second form of the equation, which is consistent with (4.2), is obtained with equal capacitor values.

If there is mismatch in the capacitor values, C_1 being $C + \frac{\Delta C}{2}$ and C_2 $C - \frac{\Delta C}{2}$, the gain will be $2 + \frac{\Delta C}{C}$ instead of 2 and the reference voltage $\left(1 + \frac{\Delta C}{C}\right) V_{REF}$ instead of V_{REF}. Similarly, a finite opamp gain will reduce the gain to $2 \cdot \left(1 - \frac{2}{A} - \frac{C_{in}}{C \cdot A}\right)$ and the reference voltage to $\left(1 - \frac{2}{A} - \frac{C_{in}}{C \cdot A}\right) V_{REF}$, where A is the opamp gain and C_{in} the capacitance from the negative opamp input node to the ground. Depending on the opamp topology, C_{in} may be much larger than the physical capacitor because of the Miller effect.

The effect of finite opamp bandwidth can be analyzed by replacing A with $A/(1 + s)$, which assumes a single pole opamp model. The bandwidth requirement is discussed in more detail later in this section.

The voltage V_{REF} is a global reference, common for all ADC stages. Thus, any deviation from the nominal value will be seen as absolute gain error, which is usually not harmful. In general, a DAC with two levels (0 and V_{REF} for instance) is always linear. In a fully differential circuit a third level can be added by defining the reference as the difference between two voltages ($V_{REF}+$ and $V_{REF}-$), to which the capacitors in the positive and negative half circuit are connected. Two levels are obtained by

alternating the polarity of the connection and the third—the zero—by shorting the capacitors together.

In higher resolution pipeline stages, more than three DAC levels are required. They can be obtained by generating more reference voltages with a resistor ladder, but then the circuit is not any more insensitive to the values of reference voltages. A better and more commonly used solution is to split the capacitor C_1 into several pieces, which can be independently connected to the references. The advantage stems from the fact that capacitor matching is typically better than resistor matching. A generic MDAC, thus, consists of a capacitor C_2, equal to C, and a capacitor C_1, which is constructed of $G-1$ pieces, each equal to C.

Several techniques for achieving resolutions higher than what is permitted by matching have been developed, from which self-calibration techniques are covered later on this section. The reference feedforward technique [66] and commutated feedback capacitor switching (CFCS) [67, 68] improve the DNL, but do not affect the INL. In 1-bit/stage architecture the capacitive error averaging technique, which has previously been used in algorithmic ADCs [69], can be used [70, 71]. With it, a virtually capacitor ratio-independent gain-of-two stage can be realized. The technique, however, requires two opamps per stage (a modification, which does not, has been proposed in [72]) and needs at least one extra clock phase.

4.5.5 Scaling

Some of the error sources in different stages, opamp gain for instance, are clearly correlated. Thus, when referred to the ADC input, they add up linearly resulting

$$V_{\varepsilon,tot} = V_{\varepsilon,1} + \frac{1}{G_1}V_{\varepsilon,2} + \frac{1}{G_1 G_2}V_{\varepsilon,3} + \cdots \tag{4.5}$$

The first stage error dominates, but when the stage gain is low, errors in the latter stages also make a significant contribution to the total error. For example, in 1.5 bits/stage architecture the total error is about $2V_\varepsilon$, when the errors in the stages are equal (V_ε). In a stage with two effective bits (gain 4) the factor 2 reduces to 1.3.

Thermal noise and random capacitor mismatch in different stages can be considered uncorrelated. Consequently, their total effect can be found by summing the squares of the input referred voltages yielding

$$V_{n,tot}^2 = V_{n,1}^2 + \left(\frac{1}{G_1}V_{n,2}\right)^2 + \left(\frac{1}{G_1 G_2}V_{n,3}\right)^2 + \cdots \tag{4.6}$$

Now the total noise in the 1.5 bits/stage architecture is only about $1.15V_n$ and practically totally determined by the first stage if stages with higher gain are used.

The stages need not be identical. In fact, considerable savings in power and silicon area can be achieved when the capacitors in the latter stages are scaled down [73], which relaxes opamp specifications and reduces the circuit area. The scaling requires budgeting a larger amount of the total noise (and capacitor mismatch) to the latter stages, which can, however, easily be done without making the first stage specifications much more stringent.

In principle, the noise contribution of every stage can be made equal, which is obtained with the scaling factor G^2. It would, however, increase the first stage specifications too much. Thus, a more reasonable scaling factor would be, for instance, G, except in the 1.5 bits/stage architecture, where even a smaller scaling factor may be optimal. Analysis in [74] shows that, from a power consumption point of view, the optimal scaling factor is between G and $G^{1.5}$; the larger the stage resolution, the larger the optimal scaling factor.

It is possible also to relax opamp gain and bandwidth in the latter stages, but not as much as capacitor size. Designing every stage with individual specifications is often too laborious. Instead, using two or three differently-sized stages is a good compromise. Scaling brings the largest advantages to the very first stages, whose requirements are the toughest. Thus, the efforts should be focused there.

4.5.6 Per-Stage Resolution

What is the best stage resolution? The smaller the stage resolution is, the larger the total number of stages becomes. On the other hand, each additional bit in a stage doubles the number of its comparators and halves the tolerable comparator offset. Furthermore, the sub-ADC input capacitance is proportional to the number of comparators. Thus, the exponential growth in the number of comparators sets an upper limit for the practical stage resolution, which seems to be around five bits.

4.5.6.1 Opamp DC Gain

When it comes to circuit speed and power consumption, the opamp is typically more important than the comparators. Thus, the effect of stage resolution on it is investigated next. The amount of sampled thermal noise is determined by the size of the sampling capacitor. To keep the noise the same regardless of the stage resolution, let the total sampling capacitance (C_{tot}) be fixed. Thus, $C_{tot} = G \cdot C$, which leads to decreasing unit capacitor C for increasing stage resolution.

The output voltage of an MDAC in the presence of finite opamp gain can be written as

$$V_{OUT} = (G \cdot V_{IN} + Q \cdot V_{REF}) \cdot \left(1 - \frac{G}{A} - \frac{GC_{in}}{C_{tot}A}\right). \tag{4.7}$$

The maximum error occurs, when the output voltage should be at its nominal maximum ($\pm V_{REF}$), which is obtained, for example, when $Q = -G + 1$ and $V_{IN} = V_{REF}$. Then the corresponding input referred error voltage is

$$V_\varepsilon = \frac{1}{G} \left(\frac{G}{A} + \frac{GC_{in}}{C_{tot}A}\right) V_{REF}, \tag{4.8}$$

which can be seen to be independent of the stage resolution [76].

Using (4.8), the required opamp DC gain in an N-bit pipeline ADC can be derived. The full-scale input range is $2V_{REF}$, making the LSB step equal to $V_{REF}/2^{N-1}$. A common requirement is that the error is smaller than the maximum quantization error, i.e. 0.5 LSB. When neglecting the input capacitance, the requirement becomes $A > 2^N$. In decibels the same is given by

$$20 \log(A) > 6.02 \cdot N \, \text{dB}. \tag{4.9}$$

In practice, the finite input capacitance slightly raises this requirement. When the effects of many stages are combined according to (4.5), a weak dependence on stage resolution is introduced, requiring, for instance, an opamp DC gain in a gain-of-2 stage 3.6 dB larger than in a gain-of-4 stage.

4.5.6.2 Capacitor Matching

Capacitor mismatch is typically proportional to the square root of capacitor size. Since the stage gain is inversely proportional to the unit capacitor size, increasing the stage resolution, which makes the unit capacitor smaller, reduces the input-referred error. Thus, a large resolution stage is less sensitive to capacitor mismatch [75].

4.5.6.3 Opamp Bandwidth

The opamp bandwidth (together with the slew rate) determines the settling time. In Appendix A the GBW requirement for an OTA in an SC amplifier has been derived. According to this, settling to N-bit accuracy within time T requires

$$GBW > \frac{(N-k) \cdot \ln 2 \cdot \left(2^k + \frac{C_{tot}}{C_L} - \frac{C_{tot}}{2^k \cdot C_L} + \frac{C_{in}\left(2^k \cdot C_L + C_{tot}\right)}{C_{tot}C_L}\right)}{2\pi T}, \tag{4.10}$$
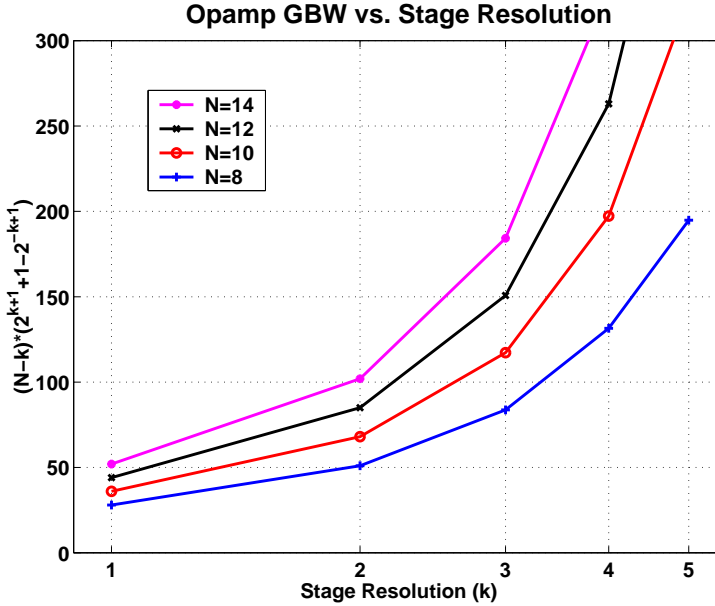
Figure 4.16 Required opamp GBW as a function of stage resolution.

where C_L is the opamp load capacitance without the MDAC capacitors. The *GBW* is defined with the OTA transconductance and the load capacitance, resulting $GBW = g_m/2\pi C_L$. Assuming that the load is primarily caused by the sampling capacitance of a similar MDAC and C_{in} is equal to $C_{tot}(2^k - 1)/2^k$ (see Appendix B), the term in parentheses becomes $2^{k+1} + 1 - 2^{-k+1}$. Thus, the required opamp bandwidth depends almost exponentially on the stage resolution. This is illustrated in Figure 4.16, where $(N-k)(2^{k+1} + 1 - 2^{-k+1})$ is plotted against k, with N values 8, 10, 12, and 14.

When the ADC sampling rate has to be maximized without capacitor scaling, 1.5 bits/stage architecture is clearly the best choice [76]. In contrast, when capacitor scaling is used and other specifications, such as power consumption and area, are also important, higher stage resolution is often preferable.

Increasing the stage gain decreases the feedback factor of the circuit, which leads to a higher opamp bandwidth requirement. Since unity gain stability is not needed from the opamp, the reduced feedback factor somewhat helps in fulfilling the increased bandwidth requirement. Higher stage gain also makes possible more aggressive capacitor scaling in the latter stages, which reduces both the total area and power consumption as well as the capacitive load to the previous stage.

Let us first consider the case where capacitor scaling is not performed. In a design space where the technology limits have not yet been reached (i.e. OTA bandwidth
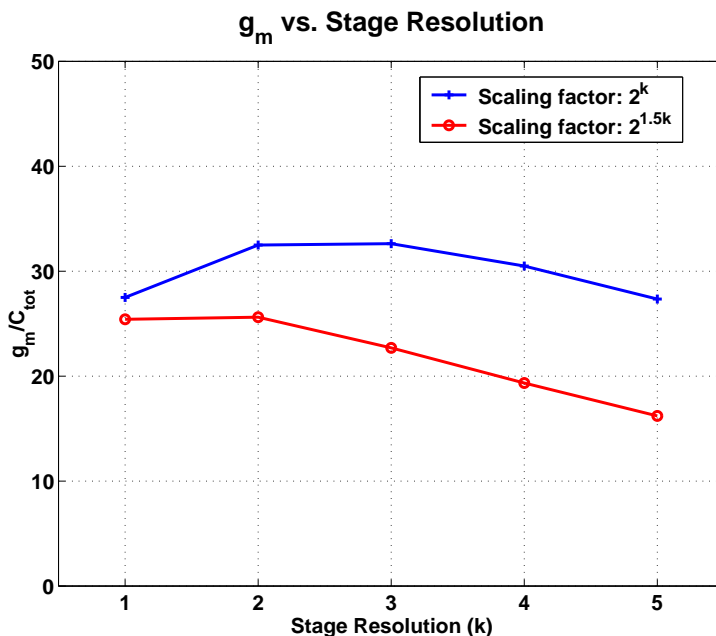
Figure 4.17 Relative opamp transconductance in a 12-bit pipelined ADC as a function of stage resolution using two different capacitor scaling factors.

can still be increased), OTA power consumption is approximately proportional to the square of g_m, on which the GBW linearly depends. When increasing stage resolution from one effective bit ($k = 1$) to two bits ($k = 2$), the GBW has to be increased by a factor of roughly 1.9, which corresponds to an increase in opamp power consumption by a factor of 3.6. Since the number of stages is simultaneously halved, the total power consumption increases by a factor of 1.8. Further increasing stage resolution makes the penalty bigger. Thus, 1-bit architecture seems also to be optimal for low power when capacitor scaling is not used. Only if opamp power consumption is dominated by the slew rate instead of the GBW requirement may a higher stage resolution be justified [4].

Let us now scale down the capacitors by a factor of 2^k between the stages, when going toward the end of the pipeline. As a result the following stage loads the MDAC with a sampling capacitance equal to $C_{tot}/2^k$. Let us assume for a moment that the opamp's load is dominated by this capacitor. The opamp *GBW* will be $2^k \cdot g_m/C_{tot}$, while the *GBW* required by the settling is proportional to $(N-k) \cdot (2^{k+2}-3)$. Thus, the requirement for the opamp g_m stays almost constant. This suggests that *maximizing stage resolution*, which minimizes the number of stages, *leads to minimum total power consumption, and the speed is rather independent of stage resolution.* More aggressive

scaling increases the advantage of high stage resolution, as seen from Figure 4.17, where the required g_m as a function of stage resolution in a 12-bit ADC is shown.

In practice, however, besides the next stage sampling capacitor the opamp load comprises its own output capacitance, the input capacitance of the next stage comparators, and the wiring capacitance as well. The part of the GBW determined by them does not improve without increasing the opamp current. Furthermore, increasing stage resolution increases the number of comparators, which partially cancels the effect of scaling. In [77] stage resolution 2 or 3 bits was found optimal for low power. The analysis, however, was made without going to the circuit level.

To summarize, it is advantageous to use high stage resolution when the sampling capacitor is large, clearly dominating the opamp load [74]. This is the situation in the front stages of a high resolution ADC [78, 79, 75]. Since 1.5 bits/stage architecture does not permit extensive scaling, a higher speed can be achieved with 2-bit or 3-bit stages.

4.5.7   Calibration

RSD logic takes care of comparator errors, but errors in the stage gain G and the reference levels produced by the sub-DACs remain in the output code. These errors are primarily dependent on capacitor matching, which is typically sufficient for 11 to 12-bit resolution. If, for some reason, the matching is worse or a higher resolution is pursued, some form of self-calibration [80] or trimming is often used.

Two approaches can be taken to calibrate out the errors: mixed signal or fully digital. In mixed signal calibration, the erroneous component values are measured from the digital output and adjusted closer to their nominal ones [81], which requires the capacitors in the MDACs to be adjustable. Alternatively, a calibration DAC can be used to sum calibration coefficients to the MDAC input [80]. Both methods apply the correction to the analog signal path, and thus require extra analog circuitry.

In the fully digital approach the component values are not adjusted [82, 83, 65]: they are just measured and used as they are. The idea behind this can be understood by looking again at the equation (4.3) for the conversion result. If the stage gain or the reference voltages deviate from their nominal values, it is not considered as an error: they are just unknown. Measuring them and using the measured values instead of the nominal ones in the formation of the ADC output code yields a perfectly correct result. The accuracy of this method depends on the accuracy of the measurement.

Pipeline architecture has been found very suitable for calibration [81, 83, 65, 84, 85, 79, 86]. The number of components to be calibrated is sufficiently small, since

only the errors in the first few stages are significant as a result of the fact that, when referred to the input, the errors in the latter stages are attenuated by the preceding gain. Furthermore, no extra ADC is necessarily required for measuring the calibration coefficients, since the back-end stages can be used for measuring the stages in front of them.

In the next equation the residue after the last pipeline stage is written on the left and the corresponding output of the final flash ADC on the right.

$$V_{IN} \cdot \prod_{i=1}^{m} G_i - \sum_{i=1}^{m} \left(\overline{Q_i} \cdot \overline{V_{R,i}} \cdot \prod_{j=i+1}^{m} G_j \right) = D_{m+1} \qquad (4.11)$$

The first term on the left is the input voltage, amplified by the total gain of the pipeline. The next term is the sum of the subtracted reference voltages, amplified by the gains of the subsequent stages. It can be seen that an error in the gain terms alone affects only the magnitude of the input voltage, which is usually not harmful. Thus, the gain terms can be left uncalibrated, which makes possible the realizing of the calibration algorithm without multipliers.

As stated earlier, only the errors in the foremost stages are significant in practice. Let us assume that only the two first stages have errors and all the following stages are error-free. The residue after the second stage equals the digital output produced by the back-end stages:

$$V_{IN} G_1 G_2 - \overline{Q_1} \cdot \overline{V_{R,1}} \cdot G_2 - \overline{Q_2} \cdot \overline{V_{R,2}} = D_{LSB}, \qquad (4.12)$$

Isolating stage 1 from the rest of the pipeline and setting the stage 2 input voltage to zero and forcing $\overline{Q_2}$ to go through all the possible bit codes, the unknown reference voltages $\overline{V_{R,2}}$ can be measured with the back-end. After that the stage 1 reference voltages $\overline{V_{R,1}} \cdot G_2$ can be measured in the same manner, the calibrated stage 2 now being a part of the back-end. It should be noted, that the calibration automatically takes into account the unideal gain (G_2) of stage 2 as well as the gain of the back-end. If more stages need to be calibrated, the procedure can be started at any point in the pipeline.

The calibration has limited accuracy, even with an error-free back-end, since the measured calibration coefficients cannot be more accurate than the resolution of the back-end ADC permits. When, during the normal operation, a term corresponding to a measured reference voltage is added to the back-end bits, the truncation error in the conversion result and in the measured value add up. In practice the number of cumulated errors is much larger, since the number of summed measurements needed

for the determining of a single reference voltage in a switched capacitor MDAC is, in worst case, proportional to 2^k, where k is the resolution of the stage [79].

To improve accuracy, the resolution of the back-end is typically enhanced by two or three bits. The calculations are performed with the enhanced resolution, after which the output is truncated to its final accuracy.

The calibration cycle to determine the coefficients is typically performed at startup. This is not always sufficient, since the component values may drift over time and change with temperature and supply voltage. Thus, the calibration has to be repeated from time to time, which requires suspending the normal operation of the converter. In many systems the input signal contains idle periods, which can be used for calibration. This is not always possible and the calibration has to be performed in the background. It can be done by having redundant circuit blocks (e.g. an extra pipeline stage [87]), so that when one element is being calibrated it is taken offline and replaced by another. Alternatively, some input signal samples can be substituted with digitally interpolated values and the freed clock cycle used for calibration [88, 79]. Running the ADC at a slightly higher clock rate than the front-end S/H circuit and queuing the sampled voltages is another way to free clock cycles for calibration [89].

If the reference levels in the MDAC are realized with resistor string instead of the more common capacitor array, the reference voltages can be measured with a separate calibration ADC in the background without disturbing the operation of the ADC [90]. Then, however, the gain error of the back-end stages is not automatically taken into account.

In [91], a dynamic element matching (DEM) DAC is used in the pipeline stage and the DAC elements are measured directly from output bitstream with a correlator during the normal operation.

4.6 Time-Interleaved ADC

Figure 4.18 shows the block diagram of an architecture in which four ADCs are used in parallel to achieve four times the sampling rate of a single converter. This is often known as time-interleaved architecture [92], since the operation of the ADC channels is interleaved in such a way that one channel processes every fourth sample. The digital outputs of the channels are combined with a multiplexer to a single full-speed bitstream.

Time interleaving was used for the first time with pipelined ADCs in 1993 in a four-channel [93] and in a two-channel ADC [94]. Since then several two-channel parallel pipeline ADCs have been published: e.g. [95, 96, 97, 98, 99, 100, 101, 102].
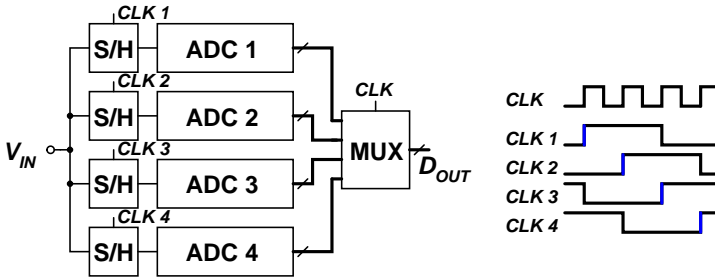
Figure 4.18 Four-channel time-interleaved ADC and its clock signals.

The implementation of a four-channel prototype, originally published in [4], will be described in Section 12.4.

4.6.1   Problems and Solutions

The problems of time-interleaved architecture arise from channel mismatch. The channels may have different offset voltages, their absolute gains can be different, or there can be a constant skew in the clock signals [92]. How these errors are seen in the spectrum of the sampled signal will be discussed in conjunction with the double-sampling technique in Chapter 9.

Up to a certain resolution, component matching is good enough and the errors can be kept to a tolerable level with careful design. High-resolution time-interleaved ADCs, however, without exception use different techniques to suppress errors.

The offset can be rather easily calibrated using a mixed signal [101] or all-digital circuitry [93, 4]. Calibrating the gain mismatch is also possible, but requires more complex circuitry than offset calibration [99, 100]. The timing skew may originate from the circuit generating the clock signals for different channels or it may be due to different propagation delays to the sampling circuits. Skew can be most easily avoided by using a full-speed front-end sample-and-hold circuit [94, 98, 4], as illustrated in Figure 4.19. The ADC channels resample the output of the S/H when it is in a steady state, and so the timing of the channels is not critical. The only problem is that the S/H circuit has to be very fast, since it operates at full speed.

The errors caused by timing skew can be corrected with digital post-processing if the signal bandwidth is limited to be somewhat smaller than the Nyquist band [103]. The technique, however, requires measuring the actual skew, even with sub-picosecond accuracy, which is not trivial. As an alternative to the digital method, the skew can be compensated with adjustable delay elements, which also requires measuring the skew.
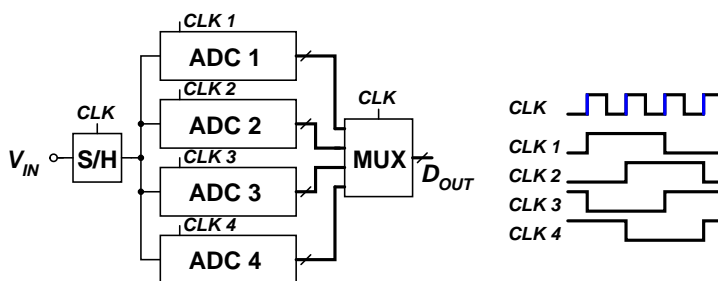
Figure 4.19 Time-interleaved ADC with front-end S/H circuit.

4.7 A/D Converters: Summary

For very low resolution (5 bits or below), flash architecture is typically the best choice. It is clearly the fastest architecture and can be scaled up to a 6–7-bit resolution range. Resolutions from 5 to 10 bits can be covered with subranging or folding and interpolating ADCs or with an architecture which is a combination of these. The speed achieved is not as high as with flash ADC, but typically somewhat higher than with pipelined ADC. Subranging and folding-and-interpolating ADCs can be realized without linear precision capacitors and the circuit area is typically quite small.

Pipelined architecture can be used in a very wide resolution range, typically from 8 to 12 bits without calibration and up to 15 bits with calibration. The interstage gain makes it possible to scale the components along the pipeline, which leads to low power consumption. In addition, the switched capacitor technique—with some modifications—has shown itself to be capable of very low-voltage operation [104, 2]. This will be demonstrated with two prototypes [8, 2] in Chapter 12.

Chapter 5

S/H Circuit Architectures

This chapter gives a brief description of the S/H architectures found in recent publications. Although the focus is on CMOS implementations, the most important bipolar architectures are also presented. The examples will show that the differences between the bipolar and the MOS device are highlighted in the design of S/H circuits. This leads to very different architectural solutions in high-performance designs.

5.1 Bipolar Architectures

5.1.1 Diode Bridge

Traditionally, the high-speed S/H circuits implemented in GaAs technology employ a diode bridge as the switching element. Since a good quality diode is often also available in silicon bipolar technology, the same type of architecture can be used. Figure 5.1 shows a simplified schematic of an S/H circuit based on the diode bridge switch. In tracking mode $\overline{clk}$ is high and the current I_1 flows through the diode bridge. Since the impedance of a forward-biased diode is very low, the output voltage across the capacitor C is almost the same as the input voltage. The circuit is turned into hold mode when $\overline{clk}$ goes low and clk goes high, steering the bias current past the bridge. The high impedance of the reverse-biased diodes virtually isolates the output from the input. In a practical circuit, at least the output has to be buffered and, usually, some additional diodes are needed to establish the reverse bias conditions.

The major disadvantage of the diode bridge S/H is the limited signal swing when operating with low supply voltages (3.3 V or less). Also, high quality diodes are not always available in bipolar technology.
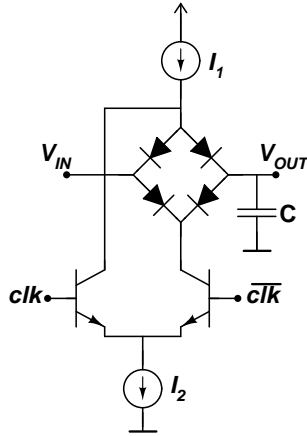
Figure 5.1 Simplified schematic of an S/H based on diode bridge .

A low voltage S/H circuit employing a single diode as a switch is reported in [105]. The circuit provides a 1.5 Vpp signal swing with a 3 V supply.

5.1.2 Switched Emitter Follower

Most bipolar S/H circuits in recent publications rely on a switched emitter follower as a switch and are usually implemented using npn transistors only. The architecture was first introduced in [106] and its schematic is redrawn in Figure 5.2. Both halves of the differential circuit consist of a switch and an output buffer. The input is brought to the switches through a differential input buffer, whose linearity is one of the major concerns in this architecture. The linearity is improved by adding emitter degeneration resistors in the input differential pair (Q1 and Q2) and diodes (Q3 and Q4) in series with the load resistors.

In track mode the buffered input is sampled in the capacitor C_H through the emitter follower Q5. In transition to hold mode the bias current of the emitter follower is turned off and its base is pulled down. The minimum size of the sampling capacitor C_H is limited by the droop rate of the held output. To make the droop smaller, the bias current of the first emitter follower (Q8) in the output buffer is turned off in hold mode. The droop in the differential output signal is considerably smaller than in the case of single output, since it is mostly common mode.

In hold mode the signal at the output of the input buffer couples to the hold capacitor through the parasitic base-emitter capacitor of Q5. The feedthrough is minimized by connecting a feed-forward capacitor C1 from the other output of the input buffer to the hold node. This capacitor is implemented by employing the base-emitter junction
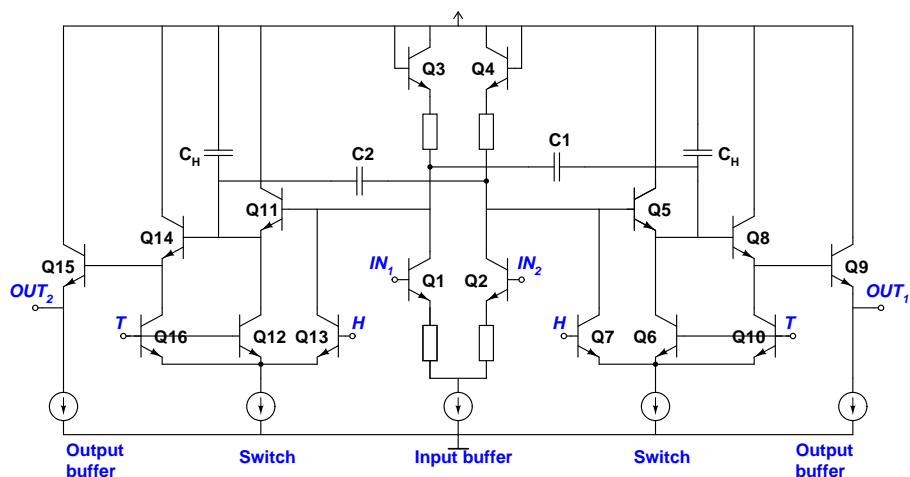
Figure 5.2 S/H architecture using emitter follower switch [106].

capacitance of a BJT.

While a 120-MS/s sampling rate was achieved in [106] it was increased to 1.2 GS/s in [107], mainly through the use of a more advanced technology. However, the increased speed was paid for by decreased linearity (from 10 bits to 8 bits) and increased power consumption (from 40 mW to 460 mW). In [108] some of the same authors managed to restore the resolution to 10 bits while maintaining the previous power consumption. This was achieved by modifying the input buffer and adding a droop compensation circuit into the output buffer. A 10-bit 250-MS/s S/H circuit with different buffers and a different hold-feedthrough cancellation technique is presented in [109]. While all the previous designs require a 5-V supply, in [110] both the input and the output buffers are redesigned so as to allow operation with a 2.7-V supply voltage.

Another low voltage (3.3 V) architecture is presented in [111]. There, the need for an input buffer is eliminated by using series type sampling, as in many CMOS architectures. The THD, measured with a 10-MHz input signal, was 60 dBc at a 100-MHz sampling rate. The quasidifferential (two single-ended circuits in parallel) circuit provides a differential signal swing as high as 3 V from a 3.3 V supply.

5.2 CMOS Architectures

One of the main challenges in bipolar S/H design is the lack of a good simple switch. The MOSFET is almost an ideal switch. When operating in the triode region it can be
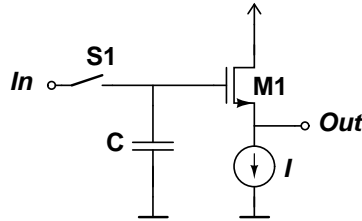
Figure 5.3 A simple S/H circuit employing a source follower buffer.

considered as a voltage-controlled resistor. The capacitive coupling over an open MOS switch is typically small and thus the hold mode feedthrough is less a problem. In addition, the purely capacitive impedance seen from the gate of a MOSFET allows the storing and buffering of sampled charges for a long time period without a significant droop.

In bipolar S/H circuits the problems are mainly connected with the linearity the buffer circuits and hold mode feedthrough via the junction capacitances of the switch device. However, by paying great attention to buffer design and using different linearization and feedthrough cancellation techniques, it is possible to make high-performance S/H circuits with open loop buffers. In CMOS the lower transconductance of the MOS device and the body effect make the simplest open loop buffer, the source follower, much less linear than its bipolar counterpart, the emitter follower.

A well-known technique used to linearize circuits is to utilize negative feedback. For example, an opamp connected to unity gain feedback makes a very linear buffer. The use of feedback does not need to be restricted to the buffer only. Enclosing the sampling capacitor in the feedback loop reduces the effects of nonlinear parasitic capacitances and signal-dependent charge injection from the MOS switches. Unfortunately, an inevitable consequence of the use of feedback is reduced speed.

The tradeoff between speed and linearity has caused researchers to take two different approaches to the design of high-speed high-resolution CMOS S/H circuits. One is to use an open loop architecture and make an effort to maximize the linearity and the other is to employ a closed loop architecture and maximize the speed.

5.2.1 S/H Circuit with Source Follower Buffer

Figure 5.3 shows a simple S/H circuit using the source follower buffer. Ideally, the channel current of a MOS transistor depends only on the gate-source voltage of the device. Consequently, a MOS transistor biased with a constant current provides a constant voltage shift from the gate to the source. The circuit has purely capacitive input
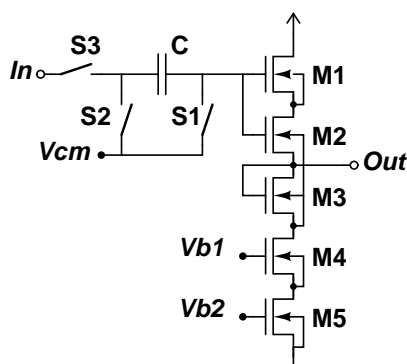
Figure 5.4 Simplified schematic of the S/H circuit presented in [112].

impedance and low output impedance, and thus it seems as if it is an ideal solution for buffering a charge stored in a capacitor. However, there are two different nonidealities that introduce input voltage dependency into the gate-source voltage. These are the bulk effect, which is the channel current dependency on the source-bulk voltage, and the finite output resistance seen when looking at the transistor from the drain. The output impedance of the current source employed in biasing the source follower also has an effect. The output impedance is inversely proportional to the channel length of the transistor and thus it has an increased importance in short channel MOSFETs.

Probably the only way to get rid of the bulk effect is to connect the source and the bulk of the transistor together. This requires that the transistor can be put in a well of its own, which is possible only with pMOS transistors in a typical CMOS process, which uses a p-type substrate. The penalty incurred by using a pMOS transistor is the slower speed in comparison to an nMOS solution. A S/H circuit employing an nMOS source follower buffer implemented in a non-typical CMOS process is presented in [112]. An implementation using a standard CMOS process and pMOS source follower is reported in [113]. The S/H circuits from [112] and [113] are shown in Figures 5.4 and 5.5 respectively.

In Figure 5.4 the effect of the finite output resistance of the source follower transistor M2 is reduced by making its drain-source voltage almost constant by cascading it with the transistor M1. To keep M2 in saturation its effective threshold voltage is made larger than the threshold voltage of M1 by biasing its bulk with the diode-connected transistor M3. The output impedance of the bias current source is enlarged by cascoding the current source transistor M5 with the transistor M4. As a result of the five stacked transistors the circuit requires a 6-V supply voltage and the signal swing is still limited to 800 mV. The circuit achieves roughly 60-dBc linearity at a 100-MHz
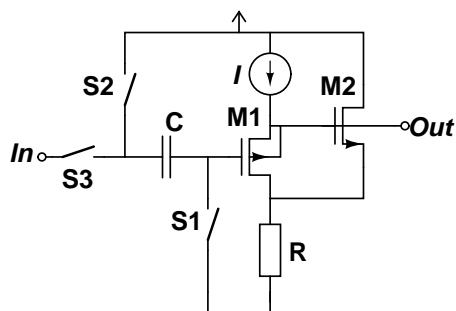
Figure 5.5 S/H utilizing a linearized source follower buffer [113].

clock rate with a 10-MHz signal frequency.

A lower supply voltage ($\sim$ 3 V) can be used with the circuit presented in Figure 5.5. The drain-source voltage of the source follower M1 is bootstrapped with a circuit consisting of M2 and R. Although this is not explained in the reference, the resistor R is used instead of a current source, probably in order to minimize the supply voltage. If a current source was used, connecting the gate of M1 through S1 to the ground would cause the current source to drop from saturation, which would slow down the transition to hold mode. To keep the current source in saturation, S1 should be connected to a higher potential, increasing the required supply voltage and circuit complexity. The drawback of using the resistor is the relatively high current flowing through M2. This is due to the fact that the bias current I determines the minimum voltage on the resistor R, which cannot be very high, in order to keep M1 in saturation when its gate is connected to the ground. The current provided by M2 should generate a voltage variation on the resistor R which is equal to the signal swing. In practice this means that the average current through M2 during hold mode can be larger than the current I. On the other hand, R must be much larger than $1/g_{m2}$ to make the drain-source voltage of M1 constant enough. It might be impossible to set the value of R so as to satisfy both these and the requirements set by the desired operating point.

In both circuits voltage-dependent charge injection from the switch transistors is avoided by taking the sample by opening the switch S1 slightly before the input switch S2. Since S1 is connected to a constant potential in both circuits, the charge it injects to the sampling capacitor is constant. Because the capacitor is floating when S2 is opened, its charge injection cannot distort the sampled voltage. This technique is called bottom plate sampling and it is discussed in more detail in Section 6.3. Since the left terminal of the capacitor is connected to a low impedance in hold mode, the attenuation to input signal feedthrough is very high.

Although distortion originating from the charge injection is prevented, a new source of distortion is introduced. Let us consider the basic S/H circuit with a source follower buffer in Figure 5.3 and assume that there is a nonlinear parasitic capacitance from the input of the buffer to the ground. It is in parallel with the sampling capacitor and thus the same voltage is sampled into both the capacitors. There is no charge redistribution in transition to hold mode and thus the nonlinear capacitance has no effect on the sampled voltage.

Let us now consider either of the circuits in Figure 5.4 or Figure 5.5. In sampling mode the input of the buffer is connected to a constant potential, while the input voltage is applied to the top plate of the sampling capacitor. In transition to hold mode the sampling capacitor is flipped over by connecting its top plate to the signal ground. The signal charge is redistributed between the sampling capacitor and the parasitic capacitor at the buffer input. Any nonlinearity in the parasitic capacitance produces harmonic distortion. The main source of nonlinear capacitance is the junction capacitance of the drain-bulk diode of the switch transistor. The capacitive loading effect of the source follower transistor is small, since both the gate-source and gate-drain voltages are almost constant. To minimize distortion the sampling switch should be small and the sampling capacitor large; thus there is a tradeoff between speed and linearity.

5.2.2 S/H Circuit Using Miller Capacitance

In [114] an interesting approach is used to reduce the signal dependent charge injection. The idea is to use the Miller effect to increase the effective capacitance in hold mode in order to render negligible the voltage step resulting from the charge injection. The sampling is fast and the switch sizes can be kept small thanks to the small physical sampling capacitor value, which is not multiplied by the Miller effect in sampling mode. The proposed circuit is shown in Figure 5.6.

The operation of the circuit is as follows. In sampling mode both the switch transistors M1 and M2 are conducting and thus the opamp is connected to unity gain feedback. The sampling capacitance is formed of the parallel combination of the capacitors C1 and C2, both connected to the low output impedance of the opamp. At the sampling instant the switch transistors M1 and M2 are turned off. The transistor M2 operates at a constant potential and thus the charge it injects into C1 does not produce distortion. The transistor M1, however, injects an input-dependent charge into node x. Now, since the feedback path around the opamp is broken, the effective value of C2 is multiplied by (A+1), where A is the open loop gain of the opamp. As a result of the increased capacitance the injected charge produces only a negligible voltage change
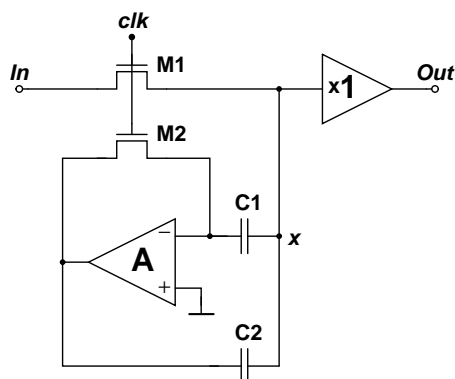
Figure 5.6 Sample-and-hold circuit using Miller hold capacitance [114].

in node x.

Since the voltage at the input of the output buffer is the same before sampling and in hold mode, the nonlinear parasitic capacitances do not distort the signal. The dominant distortion source in this circuit originates from the fact that the two switches operate simultaneously. The switches are coupled through the capacitor C1, and thus turning off M1 introduces some signal dependence to the charge injected by M2 into the capacitor C1. This phenomenon is analyzed more thoroughly in the reference.

5.2.3   Switched Transconductance S/H Architecture

A CMOS implementation of a known bipolar S/H architecture (e.g. [115]) is presented in [116]. The main idea in this architecture is to perform the sampling by turning off a MOSFET biased in the saturation region, as opposed to the more common practice of operating the transistor switch in the triode region. The advantage of biasing the switch in saturation is the fact that then the transistor channel is pinched off at the drain end. Consequently, the charge released when the transistor is turned off is injected to the source of the device, so it does not distort the sampled signal. However, since the voltage at the drain of a saturated MOSFET is not strongly defined, the switch must be enclosed in a feedback loop.

A simplified schematic of the architecture is shown in Figure 5.7. The circuit consists of an opamp, a sampling capacitor, and a unity gain output buffer. The first stage of the opamp is represented with an amplifier symbol, while the push-pull-type output stage is drawn with two transistors and a constant voltage source.

In sampling mode the feedback loop is closed and the output voltage, as well as the voltage on the sampling capacitor, follows the input voltage. The sampling is carried
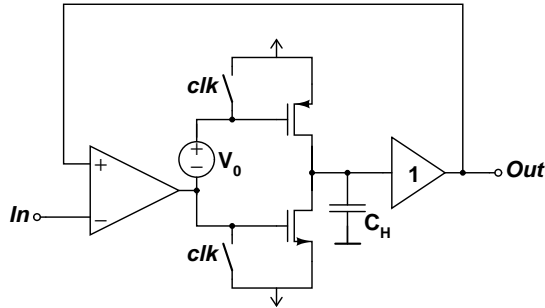
Figure 5.7 A simplified schematic of a S/H circuit using switched transconductance.
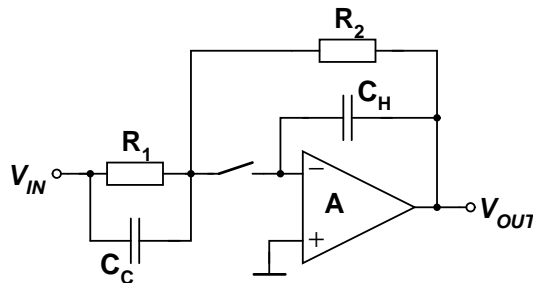
Figure 5.8 An S/H circuit whose gain is determined by resistor ratio [117].

out by turning off the output stage of the opamp by shorting the gates of the output transistors to their sources.

In hold mode the feedback loop is broken and the voltage sampled in the capacitor is buffered by the unity gain buffer. Feedback in sampling mode makes the linearity of the buffer and the capacitor irrelevant also in hold mode. Since the output transistors operate in saturation, the distortion caused by charge injection is also minimized.

In [116] the authors report 75-dBc SFDR for their pseudo-differential circuit. The closed-loop architecture limits the sampling rate to 10 kHz.

5.2.4 Closed-Loop S/H Circuit with Resistor Ratio Defined Gain

A closed-loop S/H circuit proposed in [117] is shown in Figure 5.8. In tracking mode (with the switch closed) the opamp is connected in an inverting feedback amplifier configuration and the output voltage of the circuit is $-R_2/R_1 \cdot V_{IN}$. When the switch is opened this voltage is sampled in the hold capacitor C_H. Since the switch is connected to virtual ground it does not introduce a signal-dependent charge error. The tracking mode bandwidth of the circuit is extended by adding the capacitor C_C in parallel

with the resistor R_1 to create a zero, which is used to cancel the pole due to the hold capacitor.

The circuit needs only one clock signal, which makes implementation simple. Since the circuit can realize gains other than one, it can be used as an interstage S/H in pipelined ADCs. The circuit can achieve quite high sampling rates (50 MS/s in [117] and 150 MS/s with a BiCMOS technology in [118]), but its track-and-hold nature limits the usable input signal bandwidth to far below the Nyquist frequency. Hold mode feedthrough can also become a problem, since there is a coupling path from the input to the output through the resistors. (A technique to reduce the feedthrough is proposed in [118].) Furthermore, resistor matching is known to be worse than capacitor matching in most IC technologies. Thus, it is preferable to use a circuit whose gain is determined by a capacitor ratio in applications where an accurate gain is needed.

5.2.5 S/H Circuit with Capacitor Ratio Defined Gain

When an S/H circuit with a precise gain (which is generally different from one) is needed, a switched capacitor circuit with a capacitor ratio defined gain is the best solution. Figure 5.9 shows the S/H circuit used in a pipelined ADC design [59], which is widely referred to. The input voltage is sampled passively in the capacitor(s) C_1 and in hold mode the sampled charge is transferred to the capacitor(s) C_2. The ratio of the held output voltage to the sampled input voltage is defined by C_1/C_2. In sample mode the capacitors C_2, as well as the opamp's outputs, are reset. The feedback factor of the circuit depends on the capacitor ratio and thus on the gain. The larger the gain is set, the smaller the feedback factor becomes, which increases the settling time of the circuit.

A common modification to the circuit shown in Figure 5.9 achieves faster settling by means of a modified sampling configuration. Instead of resetting the capacitor C_2 in sample mode, it is connected in parallel with C_1. Consequently, the value of C_1 has to be reduced by the value of C_2 to obtain the same gain as with the original circuit. The reduction of C_1 value increases the feedback factor in hold mode, which in turn speeds up settling. The improvement is significant only with small (~ 2) gain values.

A modification proposed in [81] adds a sampling switch between the opamp inputs. Now the original sampling switches, which are opened slightly before the new switch, need only to pass a common mode signal, which allows making them small. As a result unbalanced charge injection is reduced. Signal dependent charge injection reduces as well, since in the new configuration the voltage drop across the opamp inputs is only half of the original when the same switch size is used.
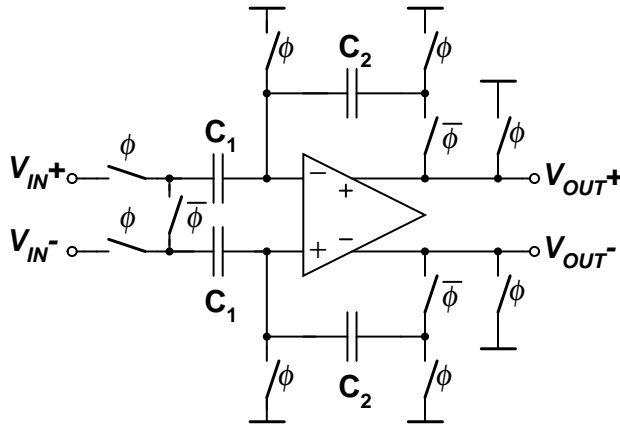
Figure 5.9 A switched capacitor S/H circuit [59].
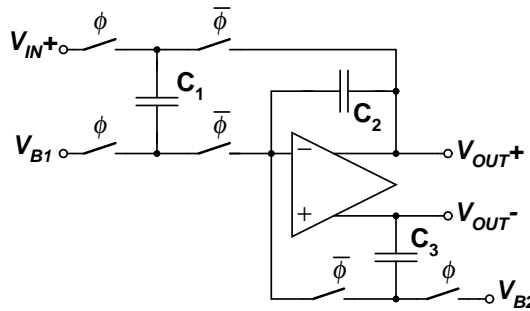
Figure 5.10 A half circuit of a fully differential S/H circuit without a reset phase [120].

5.2.6 S/H Circuit without a Reset Phase

The slew rate requirement for an S/H circuit output is set by the difference between successive output levels. Resetting the output of an S/H circuit during sample mode often makes the requirement tighter than required by the signal. A single-ended S/H circuit in which the output in the sample mode stays in the vicinity of its last hold level is presented in [119]. In [120] the idea is developed rather further by extending the length of the hold phase to the whole clock period. The fully differential circuit was later employed in [98] (the design is also reported in [121]).

The circuit from reference [120] is shown in Figure 5.10. For the sake of clarity the capacitors and switches of only one half circuit are drawn. The sampling is performed passively with the capacitor C_1 in the same way as was done in Figure 3.9. During sampling, the capacitor C_3 is connected between $V_{OUT}-$ and the bias voltage V_{B2}, which is also the opamp input common mode level. In transition to hold mode the

sampling capacitor C_1 is connected in parallel with C_2 and the bottom terminal of C_3 is switched to the opamp's negative input. Since the voltages on capacitors C_2 and C_3 are complementary, they cancel each other out when the capacitors are connected in series configuration at the beginning of hold mode. This naturally requires the capacitors to be equal in size.

Cancellation performs the reset of the hold capacitor C_2 and thus no separate reset phase is needed. At the end of hold mode the capacitor C_1 is disconnected from the feedback loop, but the output voltage remains held by the capacitor C_2. Thus the hold phase is effectively extended to overlap with the next sampling period. This, however, does not alleviate the settling requirement, since the output must be fully settled before C_1 is disconnected.

Due to the lack of a reset phase and the large feedback factor, the circuit achieves high sampling rates. A 100-MS/s sampling rate with approximately 9-bit resolution is achieved in [98].

Chapter 6

Sampling with a MOS Transistor Switch

Usually, when used as a switch, a MOS transistor is operated in the triode region (or linear region). Then the equivalent circuit for the transistor is a resistor whose value is controlled by the transistor gate voltage. When the switch is closed the value of the on-resistance is in a range from a few ohms to a few kilo-ohms. In contrast, the resistance of an open switch is so high that in practice the switch is an open circuit.

In addition to the finite on-resistance, there are also parasitic capacitances associated with the switch. This is illustrated in Figure 6.1, where a simple MOS sampling circuit is shown on the left and its equivalent RC circuit, including the parasitics, on the right. The capacitances C_{p1} and C_{p2} are due to drain and source junction capacitances and channel-to-bulk capacitance. The gate-to-source and gate-to-drain overlap capacitances and gate-to-channel capacitances are represented by the capacitors C_1 and C_2. The resistor R_{CLK} models the output impedance of the clock driver. In the
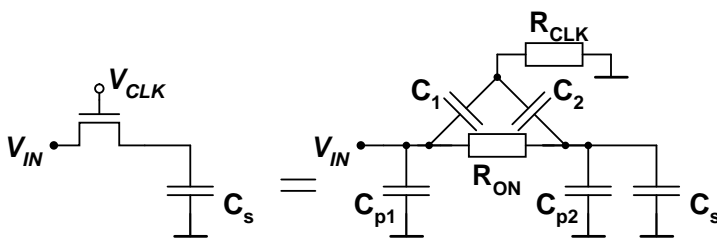
Figure 6.1 MOS sampling circuit and its RC equivalent.
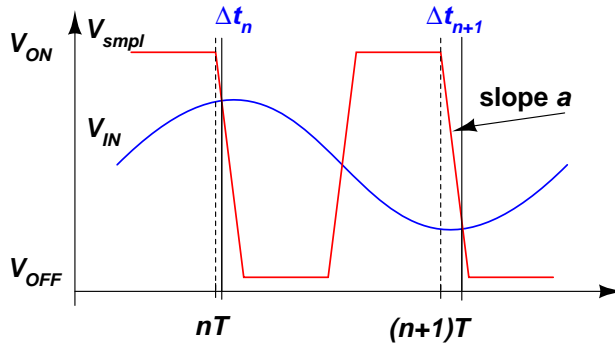
Figure 6.2 The finite slope of the sampling signal $V smpl$ results in an input voltage-dependent sampling delay Δt. For the sake of simplicity the transistor threshold voltage is assumed to be zero.

sampling circuit of Figure 6.1 the value of R_{CLK} plays an important role in hold mode feedthrough in high frequency applications, since together with C_1 and C_2, it forms a high pass coupling path past an open switch.

6.1   Voltage-Dependent Turn-Off Moment

A MOS switch like the one in Figure 6.1 turns off when its gate-source voltage becomes less than the transistor threshold voltage. (This is a simplification, accurate enough to understand the problem here. In reality the switch resistance is a continuous function of the gate-source voltage, which can be taken into account when the sampling operation is described with sampling function, as will be discussed in Section 6.5). When the switch is on, the source voltage equals the input voltage. As a result of this and the finite turn-off slope a of the gate voltage, the delay Δt from the moment when the gate voltage starts to fall to the switch turn-off moment depends on the input voltage. This is illustrated in Figure 6.2.

The following analysis shows how the voltage-dependent delay is reflected in the sampled voltage V_{OUT}. Making the assumption that the input voltage change during Δt is small, an expression for Δt can be written as

$$\Delta t \simeq \frac{V_{ON} - V_{IN}(nT)}{a} = \frac{V_{ON} - A\sin(\omega nT)}{a},\tag{6.1}$$

where the last form is obtained assuming a sinusoidal input. The output waveform can

be approximated by

$$V_{OUT}(nT) = V_{IN}(nT + \Delta t) \simeq V_{IN}(nT) + \frac{dV_{IN}}{dnT}\Delta t. \qquad (6.2)$$

For a sinusoidal input this is

$$V_{OUT}(nT) \simeq A\sin(\omega nT) + \frac{A\omega\cos(\omega nT)\left[V_{ON} - A\sin(\omega nT)\right]}{a}. \qquad (6.3)$$

Expanding the last term yields

$$V_{OUT}(nT) = A\sin(\omega nT) + \frac{A\omega V_{ON}}{a}\cos(\omega nT) - \frac{A^2\omega\sin(2\omega nT)}{2a}. \qquad (6.4)$$

From this it can be clearly seen that the input voltage-dependent turn-off moment results in second order harmonic distortion:

$$HD2 = 20\log\left(\frac{A\omega T_F}{2A_{CLK}}\right), \qquad (6.5)$$

where A_{CLK} is the clock amplitude ($V_{ON} - V_{OFF}$) and T_F the clock fall time (A_{CLK}/a). For example in a circuit where the clock amplitude is 3 V, signal amplitude 0.75 V, signal frequency 50 MHz, and clock fall time 0.1 ns, the level of the resulting second harmonic is as high as -48 dBc.

There are basically three ways to get around this problem. First, making the slope of the clock waveform steep reduces the distortion. A second and more complicated solution is to make the switch control voltage track the input signal [122]. The best solution, however, is to use a circuit topology in which the switch is operated around a constant voltage. This is discussed in more detail later in this chapter.

The bulk effect was ignored in this discussion. In practice it makes the transistor threshold voltage signal-dependent, which is another source of distortion.

6.2   Charge Injection

A conducting MOS switch has a finite amount of mobile charge in its channel. When the transistor is turned off, this charge is distributed between the source, drain, and bulk terminals of the device. To design accurate SC circuits the nature of this charge injection and redistribution phenomenon must be understood. Through the years the charge injection has been analyzed and discussed in various papers e.g. [123, 124, 125, 126].
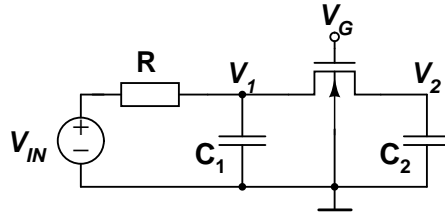
Figure 6.3 A circuit model for understanding the charge injection in MOS switches.

Consider the circuit in Figure 6.3. There, a sampling capacitor is driven by the voltage source V_{IN} with the source resistance R through a nMOS switch transistor. The capacitances C_1 and C_2 are associated with the source and drain terminals of the transistor, C_2 including the sampling capacitor. The distribution of the charge is dependent on the ratio of C_1 and C_2 as well as the source resistance R and the waveform of the switch control voltage V_G. The total amount of the channel inversion layer charge is dependent on the voltage V_{IN}.

The amount of the released charge Q_{tot} can be expressed [126] as

$$Q_{tot} = C_G(V_{GON} - V_T), \tag{6.6}$$

where C_G is the total gate channel capacitance, V_{GON} the transistor gate voltage in the on-state and V_T the effective threshold voltage. A first order approximation is that C_G is not dependent on the input voltage and V_T has a linear input voltage dependency through the bulk effect. In that case the amount of released charge is linearly dependent on the input voltage, which is experimentally verified in [123] and [124]. To model the charge more accurately, the nonlinear bulk effect and voltage dependency of C_G have to be taken into account.

As the transistor is turned off, a part of the inversion layer charge is leaked to the substrate. This phenomenon, known as charge pumping, is due to two effects, the capture of charge by the interface traps and recombination in the channel and the substrate. It is shown in [126] that substrate leakage occurs only when the gate voltage turn-off slope is extremely steep or the transistor channel is very long, and thus in practical switches this effect can be ignored.

The remaining question is how the inversion layer charge is distributed between the drain and the source terminals when the switch is turned off. This can be analyzed by using the circuit shown in Figure 6.4 to model the distributed gate capacitance and assuming the gate voltage to have the waveform shown in Figure 6.5 [126]. When the switch is conducting the gate voltage has a value of V_{GON} and when it turned off the
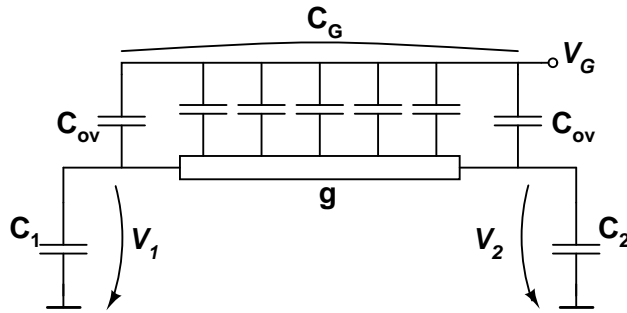
Figure 6.4 A circuit model for distributed gate-channel capacitance.
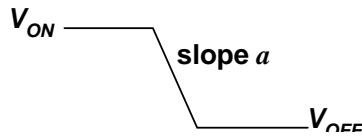
Figure 6.5 The gate voltage turn-off waveform.
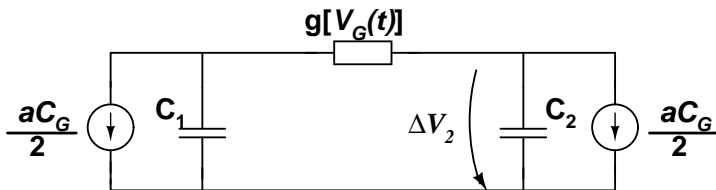
Figure 6.6 A simplified MOS switch model for charge injection analysis.

voltage is switched to V_{GOFF} with the slope a. Using this, the channel conductance g can be written as

$$g[V_G(t)] = \beta(V_G(t) - V_T) = \beta(V_{GON} - at - V_T), \quad (6.7)$$

where $\beta = (W/L)\mu C_{OX}$. Now, the transistor channel during the turn-off can be modeled with the time-varying conductance g. The injected charge is modeled with current sources having a value of $aC_G/2$ in parallel with the capacitors C_1 and C_2, as shown in Figure 6.6. The analysis of this circuit yields a differential equation

$$\frac{dV}{dT} = (T - B)\left[\left(1 + \frac{C_2}{C_1}\right)V + 2T\frac{C_2}{C_1}\right] - 1, \quad (6.8)$$

where the normalized factors are

$$V = \frac{\Delta V_2}{\frac{C_G}{2}\sqrt{\frac{a}{\beta C_2}}} \quad (6.9)$$

$$T = \frac{t}{\sqrt{\frac{C_2}{a\beta}}} \quad (6.10)$$

$$B = (V_{GON} - V_T)\sqrt{\frac{\beta}{aC_2}} \quad (6.11)$$

Solving the equation (6.8) numerically gives the diagram shown in Figure 6.7. There, the quantity $\Delta Q_2/Q_{tot}$ is expressed as a function of parameter B. The family of curves represents the solutions with different C_1/C_2 ratios. When the values of B are small the total charge is equipartitioned between the two capacitors, regardless of the capacitor ratio. On the other hand, when B is large, the charge is partitioned according to the capacitor ratio. The intermediate B values result in a partitioning somewhere between the extreme cases.

Since the parameter B is dependent on the turn-off time through the slope a, the meaning of the result can be understood as follows. When the transistor is turned off rapidly, the channel is cut off before the potential difference between drain and source has time to even out and, as a result, the channel charge is equally divided between the drain and source terminals. On the other hand, a slow turn-off leaves time for the drain and source voltages to become equalized, which results in a charge partitioning according to the capacitor ratio. In this analysis the source resistance R is assumed to be infinite. A more complete study with finite R values is performed in [125] showing
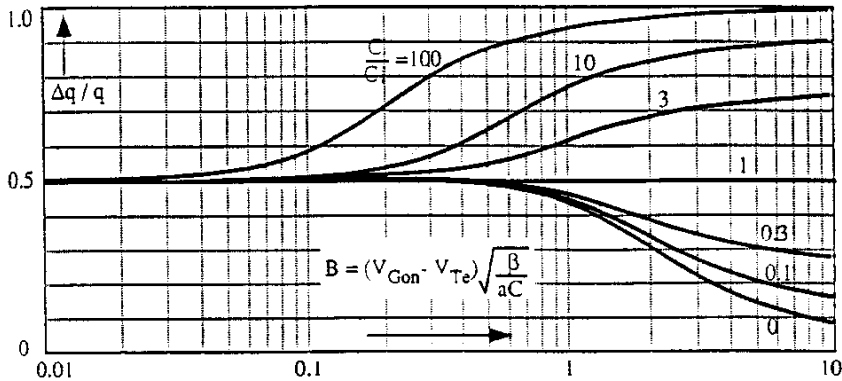
Figure 6.7 Charge partitioning as a function of *B* [37].
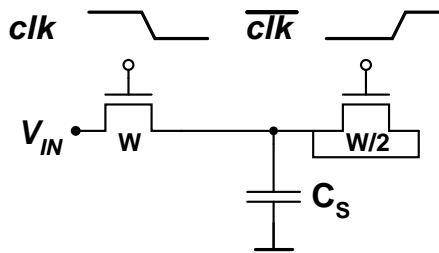
Figure 6.8 Canceling the charge injection with a dummy switch.

that the smaller the source impedance is, the smaller the part of the injected charge that ends up in the capacitor C_2 is.

In practical circuits the slope of the gate voltage is usually in the region where the charge partitioning is dependent on the slope and capacitor values. Consequently, the amount of injected charge is not well controlled and as a result of this, different strategies are used and proposed to overcome the problem. First, the capacitor C_1 in parallel with the driving voltage source and the parameter B can be made large, so that the injected charge returns to the driving circuit. This, however, increases the capacitive load and makes the circuit slow. Another strategy is to make the capacitors equal so as to make the charge injection equal as well. The same result can be achieved by making the parameter B much smaller than 1. In the latter two cases, the charge injection can be canceled by using a half-sized dummy switch as illustrated in Figure 6.8.

The first order approximation of the amount of injected charge (6.6) indicates that the charge is linearly dependent on the input voltage. In the S/H circuit shown in

Figure 6.1 this only affects the gain of the circuit, which is usually not very harmful. In practice, the charge injection also has a nonlinear component, which results in harmonic distortion. Even the linear component alone can cause distortion in some types of SC circuits, e.g. by changing the interstage gain in pipelined ADCs.

When simulating SC circuits in SPICE the designer should be aware that the charge injection is not completely modeled in older MOS models. The quasi-static transient model, used in those, may give incorrect result, especially when the slope of the gate voltage is very steep [127]. In more recent models, such as BSIM3v3, more accurate charge injection modeling is achieved by employing a non-quasi-static model.

6.3   Bottom Plate Sampling

The discussion in the previous sections has shown that both the signal-dependent charge injection and the signal-dependent turn-off moment originate from the fact that the switch transistor sees the input voltage in its source terminal. If the switch was operated around a fixed voltage, the error resulting from both phenomena would be constant. A constant error is less harmful in many applications and it can be reduced with a differential circuit topology.

In many closed-loop S/H architectures the sampling switch is connected to a virtual ground to avoid signal-dependent errors. An example is shown in Figure 6.9. The feedback loop includes two opamps, which inevitably slows down the circuit. By using more than one switch, sampling against a constant potential can be achieved without enclosing the switch in the feedback loop. This well-known technique [128], called bottom plate sampling (also known as series sampling), is illustrated in Figure 6.10. The capacitor C is the sampling capacitor and the capacitor C_L is a parallel combination of the input capacitance of the following circuitry and the parasitic capacitances associated with the switch S_2 and the sampling capacitor.

The idea goes as follows. In sampling mode the switches S_1 and S_2 are conducting, while the switch S_3 is open; thus, the input voltage is sampled in the capacitor C. At the sampling instant the signal $\phi\prime$ goes down, opening S_2, which leaves node V_{OUT} floating. Since the switch S_2 is always connected to the ground the charge it injects into node V_{OUT} is constant. Slightly later, the capacitor C is disconnected from the input by opening the switch S_1. The charge injection and the input voltage variation during the time gap between opening S_2 and S_1 distort the voltage on C. This, however, is not dangerous, since the sampled signal is in the form of a charge at node V_{OUT}. This charge cannot change after S_2 is opened, because there is no other DC path from
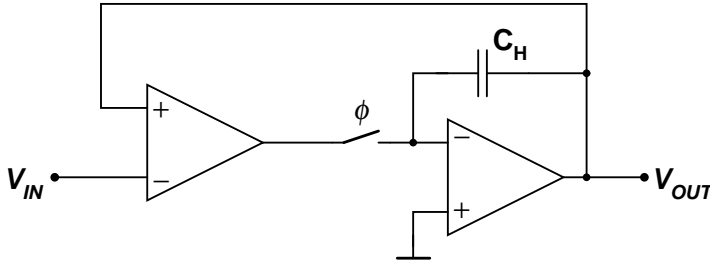
Figure 6.9 An example of a closed-loop architecture where the sampling switch operates against a fixed voltage.
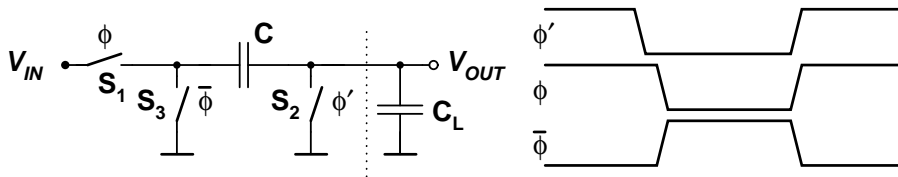
Figure 6.10 Bottom plate sampling.

that node. The sampling is completed by connecting the left-hand terminal of the capacitor C to the ground by closing the switch S_3.

In hold mode the signal can be taken out from node V_{OUT} either as a voltage or as a charge. If V_{OUT} is connected to a high impedance its voltage is just an inversion of the sampled input voltage. In practice, the load capacitance C_L causes some attenuation and, if the capacitance is signal-dependent, harmonic distortion. Alternatively, node V_{OUT} can be connected to a virtual ground, which makes it possible to handle the sample in the form of a charge.

6.4   Nonlinear Time Constant

In sampling mode the circuit in Figure 6.1 forms a low pass filter; consequently, the maximum frequency that the circuit can track is limited. The 3dB frequency of the circuit is

$$f_{3dB} = \frac{1}{2\pi R_{ON}(C_S + C_{p2} + C_2)}, \tag{6.12}$$

where the clock driver output impedance R_{CLK} is assumed to be zero. If absolute accuracy is required then the minimum 3dB frequency is given by [10]

$$f_{3dB} > 2^{(N-1)/2}f, \tag{6.13}$$

where f is the frequency of the input signal and N the resolution in bits. Using these two equations the maximum switch on-resistance can be calculated. For example, 12-bit accuracy in the 100-MHz signal band sets R_{ON} to 18 Ω, when a total capacitance of 2 pF is assumed. This begins to approach the limits of what is achievable with a single-transistor switch or a transmission gate with today's technologies. In most circuits, however, some attenuation can be tolerated, making higher resolutions and wider bandwidths possible.

Unlike attenuation, harmonic distortion is intolerable in most applications. When the signal amplitudes are large, accuracy and signal bandwidth are limited by distortion, which originates from the fact that switch on-resistance and stray capacitances are not constant but vary as functions of drain and source voltages. For a short channel device the on-resistance is

$$R_{ON} = \frac{1 + \frac{V_D - V_S}{E_c L}}{C_{ox}\mu_{eff}\frac{W}{L}\left[V_G - \frac{V_S}{2} - \frac{V_D}{2} - V_{TO} - \gamma\left(\sqrt{V_S - V_B - 2\phi_F} - \sqrt{2\phi_F}\right)\right]}, \qquad (6.14)$$

where V_G, V_S, V_D, and V_B are the voltages on the transistor's gate, source, drain, and bulk terminals. By looking at the equation three different signal-dependent terms can be identified. The first and clearly dominant one is the gate-channel voltage $V_G - (V_S + V_D)/2$ in the denominator. The second is the threshold voltage dependency on the source-bulk voltage (bulk effect) modeled with the square root terms in the denominator. The last is the term in the numerator which depends on the drain-source voltage, the critical electric field E_c, and the device channel length.

The dominant nonlinear parasitic capacitances are the drain and source junction capacitances, which are given by [129]

$$C_{jbx} = C_{j0} \cdot \left(1 - \frac{V_B - V_X}{P_p}\right)^{-Mj}, \qquad (6.15)$$

where V_X is the drain or source voltage, C_{j0} the junction capacitance with a zero bias, P_p the bulk junction potential, and Mj the bulk junction grading coefficient. Actually, the capacitance is a sum of two components, the tub bottom capacitance and the sidewall capacitance, which both follow equation (6.15) but with different parameters.

There are basically two ways to reduce distortion: decreasing the absolute value of the time constant and making the time constant less nonlinear. The remainder of this section discusses these techniques in more detail, reviews known switch circuits, and proposes some new ones.
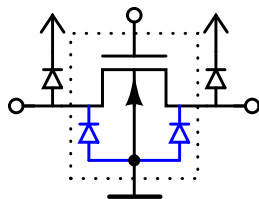
Figure 6.11 The junction capacitances of an nMOS transistor are linearized with diodes.

6.4.1 Linearization of Basic Switches

The commonly used basic switch—the transmission gate—can itself be considered as a linearized circuit; as the signal voltage rises the increase in the on-resistance of the nMOS transistor is compensated for by the decrease in pMOS on-resistance and vice versa. Similarly, as the voltage rises, the drain and source junction capacitances of the nMOS decrease, while in the pMOS the opposite happens.

The relative size of the transistors (the width of the pMOS compared to the width of the nMOS) can be optimized in order to minimize distortion, as demonstrated with simulations by the author in [10]. In principle, to the first order (ignoring the junction capacitances, the bulk effect, and the short channel effects) the time constant can be made flat in the voltage range where both the transistors are conducting (from $V_{T,p}$ to $V_{DD} - V_{T,n}$). Optimum sizing, however, is rather sensitive to process parameters and thus is not the same in different process corners. Consequently, size optimization can yield only moderate linearity improvements.

In a single-transistor switch the junction capacitances can be linearized by putting diodes of opposite type in parallel with the junctions [130], as shown in Figure 6.11. This makes the capacitances more symmetrical about the mid-supply level, which decreases the even order distortion.

The DC level of the signal has a large impact on the harmonic distortion, which is especially emphasized with single-transistor switches. When using nMOS switches it is advantageous to situate the signal range as low as possible, since switch on-resistance and distortion increase rapidly toward high voltage values. In low-voltage designs, however, the signal range is usually a considerable portion of the supply voltage and so the DC level cannot be set far apart from the mid-supply level.

Some technologies offer low-threshold devices, which can be used to extend the signal range and reduce distortion. In very low-voltage circuits the gained extra gate overdrive of some hundreds of millivolts can help a great deal. The low threshold voltage, however, may prevent the switch from being properly turned off, which can
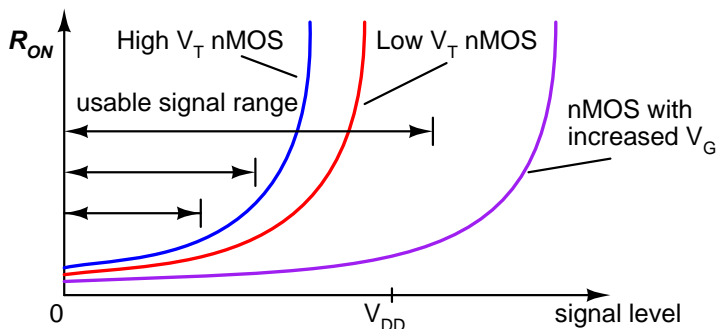
Figure 6.12 nMOS switch on-resistance as a function of signal level.

result in a leakage of the stored charge.

An effective way to reduce switch on-resistance and to extend the linear range is to employ a voltage higher (in the case of nMOS) than the supply used to control the switch transistor gate. This is illustrated in Figure 6.12, where the on-resistances of nMOS switches with different gate overdrives are plotted as a function of signal level. There are couple of ways to realize the higher voltage. The most straightforward is to supply the voltage externally, which, however, is often too costly from the system point of view. Thus, a better solution is to generate the voltage on chip. Since the current drain from this supply is relatively small, it can be easily implemented with a charge pump. To avoid potential cross talk problems and routing the voltage around the chip, the voltage generation is often distributed. When this principle is adhered to as far as possible, each switch has a charge pump of its own.

6.4.2   Gate Voltage Boosting

A switch with a local charge pump circuit is shown in Figure 6.13 [23]. There, the capacitor C1 is charged to $V_{DD} - V_T$ when *clk* is high. At the same time, the gate of the switch transistor is held in the ground by the transistor M3. When the clock goes down C1 boosts the gate of the switch transistor to $2V_{DD} - V_T$. In practice, the voltage is somewhat lower because of parasitic capacitances.

Another local charge pump circuit is shown in Figure 6.14 [73, 131]. In the previous circuit the capacitor precharging is carried out through the diode-connected nMOS transistor M1. This diode switch limits the precharge voltage to $V_{DD} - V_T$. In Figure 6.14 the capacitor (now C2) is precharged through an nMOS switch M2. The capacitor can be charged to V_{DD} since the gate of M2 is controlled by a boosted voltage generated with M1 and C1 [132]. The well bias for the pMOS transistor M3 is
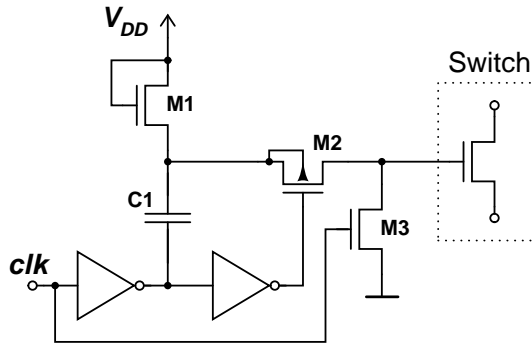
Figure 6.13 A MOS switch with local gate voltage boost circuit [23].
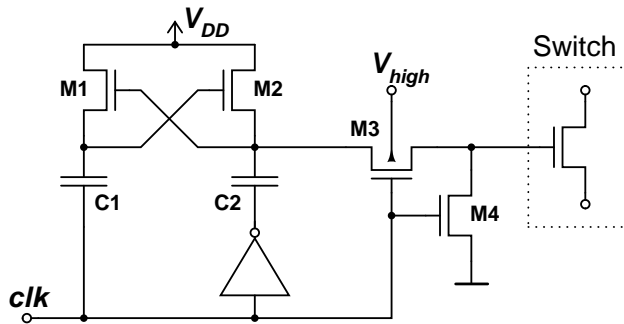
Figure 6.14 Another local boost circuit for a MOS switch [73].

produced with another charge pump. More switch boost circuits can be found from references [122, 133, 134].

6.4.3   Bootstrapped Switches

The boosted switches presented above are not generally suitable for deep submicron technologies, because the high voltage levels cause reliability problems. Four mechanisms which can degrade a MOS device over time or suddenly destroy it are gate oxide breakdown, gate-induced drain leakage, hot-electron effects, and punch-through [104]. The first one is the result of too-high gate-source voltage, the second of too-high gate-drain voltage, and the last two of too-high drain-source voltage.

The nominal supply voltage of a deep submicron technology is typically set as high as reliability permits. Consequently, in a long-term reliable circuit the terminal voltages of a transistor may not be much higher than the supply. There are, however, two things that should be noticed. First, the voltages are not absolute values but rela-
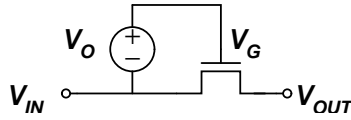
Figure 6.15 A switch controlled with input tracking gate voltage.
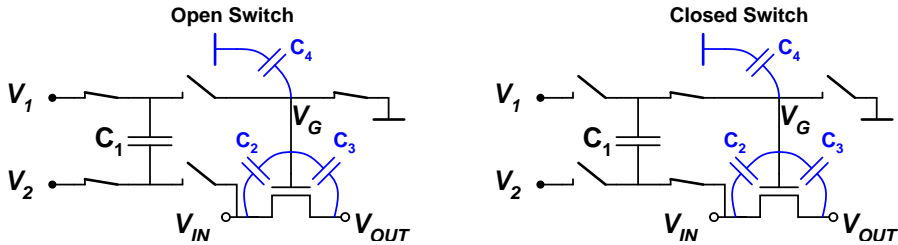
Figure 6.16 SC implementation of the offset voltage source.

tive voltages between the device terminals, and second, stress accumulates over time and so the device's life-time depends on the average (or DC) voltage rather than on the maximum value. Of course, a sufficiently high voltage can cause instantaneous failure.

6.4.3.1 Principle

To avoid violating the technology reliability specifications, the transistor gate-source (or gate-drain) voltage may not exceed the nominal supply voltage. Taking this condition into account, switch boosting can still be realized by making the gate voltage track the source voltage with an offset V_O, which is, at its maximum, equal to the supply voltage. This technique, called bootstrapping, is illustrated in Figure 6.15, where a closed switch with a gate voltage $V_{IN} + V_O$ is shown. Besides reliability, another advantage of this circuit is that now the switch gate-source voltage is constant and, as a result, a major source of nonlinearity in equation (6.14) is greatly attenuated.

The voltage source in Figure 6.15 can be implemented with a switched capacitor, which is precharged in every clock cycle. Such a circuit is shown in Figure 6.16. During the clock phase when the transistor is non-conductive the capacitor C_1 is precharged to $V_1 - V_2$. To turn the switch on, the capacitor is switched between the input voltage and the transistor gate. The gate voltage, however, is not exactly the sum of the input voltage and the precharge voltage, since the parasitic capacitances associated with the switch transistor and the auxiliary switches cause some distortion.
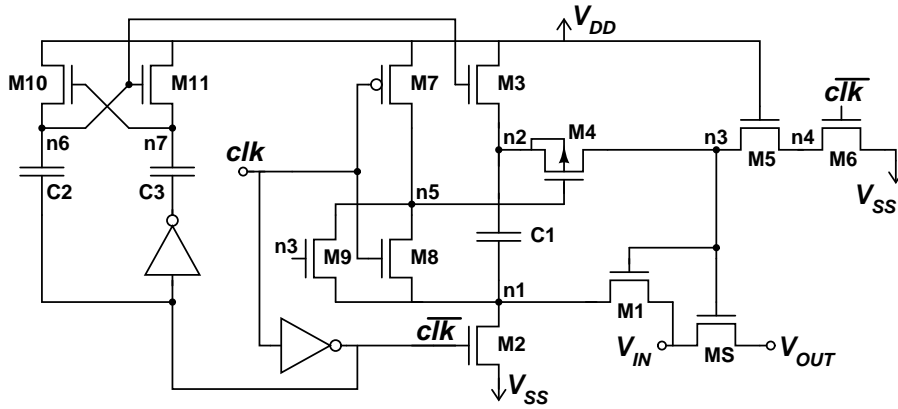
Figure 6.17 A long-term reliable bootstrapped switch [135, 104].

The gate voltage is given by

$$
\begin{aligned}
V_G &= \frac{C_1(V_1 - V_2)}{C_{tot}} + \frac{(C_1 + C_2 + C_3)V_{IN}(t)}{C_{tot}} \\
&\quad - \frac{C_2 V_{IN}(t_0)}{C_{tot}} - \frac{C_3 V_{OUT}(t_0)}{C_{tot}} \\
&= V_O + V_{IN}(t) - \frac{C_4 V_{IN}(t)}{C_{tot}} - \frac{C_2 V_{IN}(t_0)}{C_{tot}} - \frac{C_3 V_{OUT}(t_0)}{C_{tot}},
\end{aligned}
$$

(6.16)

(6.17)

where C_{tot} is the total capacitance $(C_1 + C_2 + C_3 + C_4)$. In the second form the first two terms—the offset voltage V_O and the input voltage—are the desired ones, while the three others are unwanted. The capacitive division between the capacitances from the input node to the gate and from the gate to the signal ground results in a term proportional to C_4. The last two terms, proportional to the drain and source overlap capacitances of the switch transistor, result from the fact that at the end of the switch off-phase (time t_0) the voltages at the source and at the drain are sampled into the parasitic capacitances. In order to minimize distortion, the parasitic capacitances have to be minimized and the bootstrapping capacitor C_1 made large.

6.4.3.2 Circuits from the Literature

A practical implementation of the bootstrapped switch is shown in Figure 6.17 [135, 104]. There, the offset voltage is realized with the capacitor C1, which is precharged to V_{DD} during the main switch off-period. To turn on the switch MS, the precharged capacitor is connected between its source and gate via the series switches M1 and M4. Turning off MS is performed by disconnecting the capacitor C1 and pulling down the
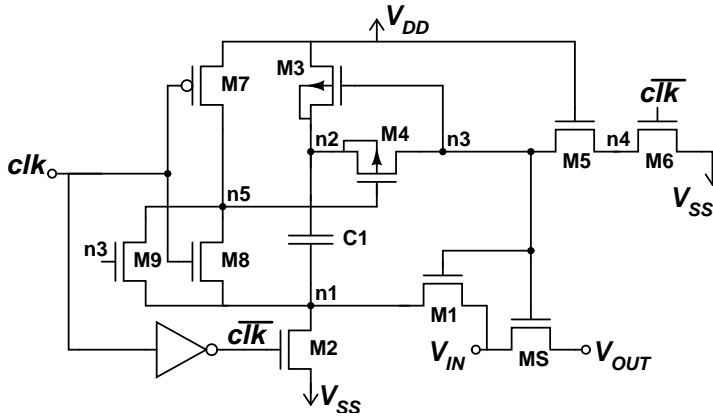
Figure 6.18 Another bootstrapped switch [136].

gate with M6. The transistor M5 is needed to prevent the gate-source voltage of M6 from exceeding V_{DD}.

A slightly modified version of this circuit is presented in [136] and shown in Figure 6.18. There the nMOS transistor M3 is replaced with a pMOS transistor whose gate is tied to the gate of MS. Consequently, the charge pump on the left in Figure 6.17 can be eliminated.

Some other bootstrapped switches are presented in [137], [138], and [139]. These circuits, however, are not targeted on deep submicron technologies and thus not necessarily suitable for low supply voltages. In baseband delta-sigma modulators the signal does not change much within a clock cycle, making it possible to bootstrap the switch with a signal value sampled half a clock period earlier [140, 141].

6.4.3.3 Eliminating the Bulk Effect

The linearity of these circuits is limited by second order effects, the two most important of which are the bulk effect and the voltage dependent junction capacitances. In a triple well process, an example of which is a typical BiCMOS processes, both of them can be almost completely eliminated by connecting the switch transistor bulk to the input node during the tracking phase. In the circuit shown in Figure 6.18 this can simply be done by connecting the bulks of transistors MS and M1 to node n1. They cannot be connected directly to the input, because when the switch is open the voltage V_{IN} may go far below V_{OUT}, making the diode from the output node to the bulk forward biased. In some circuits a buffer [138, 139] or a resistive circuit [142] is inserted between the input and the bulk node to isolate the capacitance of the well-to-substrate
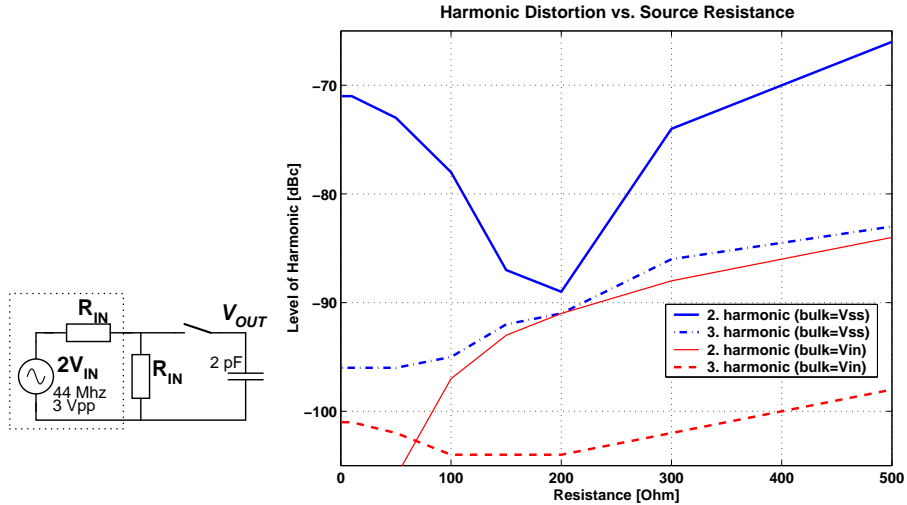
Figure 6.19 Simulated distortion as a function of signal source output impedance.

junction from the input. This, however, increases the circuit's complexity and introduces phase shift between the input and the bulk voltage. Thus, the advantage gained is questionable.

In some cases the elimination of the junction capacitances worsens the distortion, because normally the capacitance nonlinearity partially cancels the on-resistance nonlinearity; for instance, in an nMOS device the junction capacitance decreases, while the on-resistance increases, as the voltage level rises. Which source of nonlinearity dominates depends on the output impedance of the signal source.

This is investigated with a simulation, in which the levels of the second and the third harmonic in the tracked signal are calculated with different values of signal source output impedance. The simulated circuit is a simple SC sampling circuit utilizing the switch presented in Figure 6.18, with the bulks of MS and M1 connected to n1. The simulation method used is transient analysis. For comparison the simulation is also performed for the same circuit with the bulks of transistors MS and M1 connected to V_{SS}. The signal source is a 44-MHz, 3.0-Vpp sinusoidal voltage source with a series resistance R_{IN}. An equal termination resistor is put in parallel with the sampling circuit, resulting in a 1.5-Vpp input signal for the switch.

The results are shown in Figure 6.19, which proves that the level of the second harmonic has a strong dependency on the resistance. Moreover, in the switch with bulk node connected to a constant potential the two nonlinearity sources partially cancel each other out, resulting in a distortion minimum at 200 Ω. In the third harmonic,

Figure 6.20 A bootstrapped switch without bulk effect in a standard CMOS technology.

which is more important in fully differential circuits, a similar clear cancellation cannot be seen. Despite the cancellation, the circuit with input tracking bulk yields lower distortion over the whole resistance range; thus, it should be used whenever the technology permits.

The improvement in linearity achieved by making the bulk node track the input is so remarkable that accomplishing the same with a standard CMOS technology would be desirable. The problem there is that the bulk node of the nMOS device (in a typical process with p-type substrate) cannot be accessed. Thus, the switch transistor has to be implemented with an inherently more resistive pMOS device.

Let us consider the circuit shown in Figure 6.18 and change every nMOS transistor to pMOS and vice versa. Most of them can be changed without problems, but not the transistors M3 and M4, because now nodes n2 and n3 go below V_{SS}, which would make the junction diode of any nMOS device connected to these nodes forward biased.

Author's proposal as to how to overcome this problem is shown in Figure 6.20. There, the device M3 is implemented with a diode-connected pMOS device as in Figure 6.13, but it can also be implemented as in Figure 6.17.

The implementation of the series switch between the gate of the transistor MS and the bootstrap capacitor is more difficult. The switch has to be a pMOS transistor with the ability to conduct at voltage levels much below V_{SS}. Thus, its gate voltage has to be bootstrapped.

Figure 6.21 Simulated waveforms in the circuit shown in Figure 6.20.

The operation of the circuit can best be understood by looking at the simulated voltages in Figure 6.21. When the main switch is turned off, its gate node n3 is shorted to V_{DD} and the gate of M4 is connected to n3 via M7. The off-phase ends when *clk* goes up. First, M7 is turned off and M8 on, and, as a result, the voltage at node n5 starts to go down. Since n3 is still shorted to V_{DD}, a voltage appears across the capacitor C2. A short delay later, *clk2* also goes up, releasing node n3, which now starts to go down as M4 opens. When the voltage at n5 reaches the threshold level of M8, the transistor automatically turns off, leaving node n5 floating. The bootstrap capacitor C2 makes n5 follow n3 with an offset large enough to keep M4 properly conducting. To lower the threshold voltage of M4 its bulk is switched to V_{SS} during the on-phase.

To ensure that the circuit is reliable in the long term it has to be made sure that the gate-source or gate-drain voltage of the transistors connected to nodes n2, n3, and n5 will not exceed V_{DD}. Since the gate voltage of M4 follows node n3 there are no problems with that device. The reliability of M7 and M8 is guaranteed by connecting their gates to n4 instead of V_{DD}, when it is desired that the devices are off.

Since the transistor M4 is not a very large device and its on-resistance does not need to be very linear, the second bootstrap capacitor C2 and the transistors M7 and M8 can be fairly small. Thus, the added parasitic capacitance is not large.

The linearity of the proposed circuit has been compared with the circuit shown in Figure 6.18 with transient simulations. The results show a lower third harmonic for the proposed circuit at signal source resistance values greater than 250 Ω and a lower second harmonic at all resistance values except the optimum 200 Ω of the reference circuit (see Figure 6.19). In conclusion, the circuit is the preferable choice in single-ended applications with a low to moderate impedance signal source (e.g. 50 Ω) and in all applications with a high impedance signal source (>250 Ω).

Two other standard CMOS circuits for eliminating the bulk effect have been proposed in the literature. In the first one [143] the bootstrap capacitor is connected directly, without a series switch, to the gate of the switch transistor. During the off-state this capacitor terminal is shorted to V_{DD} and the other terminal is charged to $2V_{DD}$ with an additional charge pump.

The second circuit [144] removes the bulk effect, but not the effect of the junction capacitances. There, the signal voltage is not directly connected to the gate of the switch device, but is first predistorted to account for the bulk effect. This is done by generating a proper gate-source voltage with a dummy switch, whose source tracks a buffered signal voltage. The dummy switch operates in the saturation region and it carries a constant current.

6.4.3.4 Double-Side Bootstrapping

By looking at the on-resistance equation (6.14) it can be seen that the dominant signal-dependent term $V_G - (V_S + V_D)/2$ is not made exactly constant with the circuits presented thus far, since, although the drain and source voltages are close to each other, they are not the same because of the signal current through the switch; thus, the achievable linearity is limited.

A proposal as to how to overcome this problem is shown in Figure 6.22. There, the idea is to make the switch transistor gate follow the average of the drain and the source voltage rather than either of them alone. This is accomplished by adding another bootstrap capacitor on the output side of the switch transistor, and thus a suitable name for the technique is double-side bootstrapping. The total capacitance need not be increased, since the original bootstrap capacitor can be divided into two equal pieces without sacrificing accuracy. It appears from simulations that the improvement achieved in linearity can be of the order of 5 decibels on the level of the 2nd harmonic.

A slightly different double-side bootstrapped switch is proposed in [145], however, not for enhancing linearity, but for improving reliability at the beginning of the on-phase when it is not known which side of the switch transistor is in lower potential.
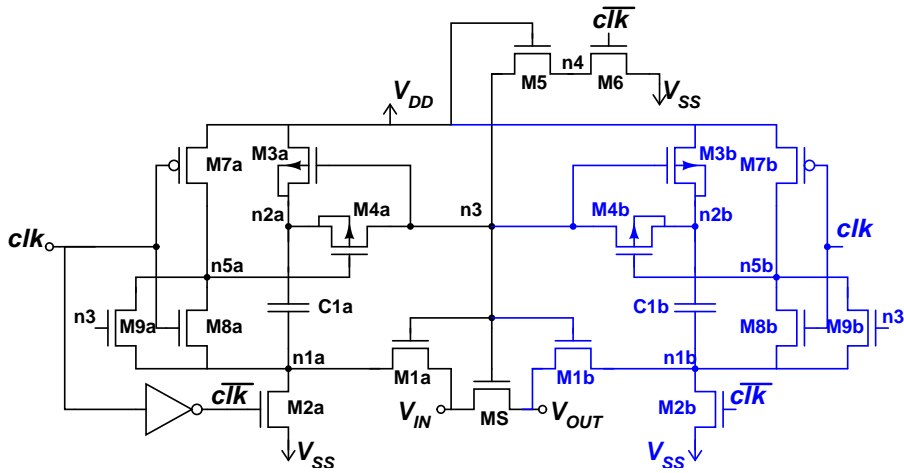
Figure 6.22 Double-side bootstrapped switch.
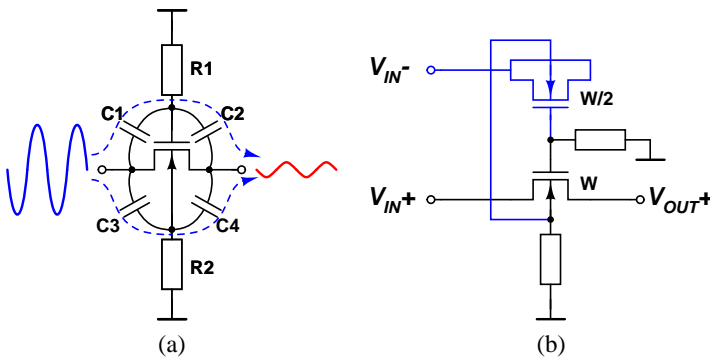
Figure 6.23 High-pass feedthrough paths past a closed switch and a technique to cancel the feedthrough.

6.4.3.5 Reducing Feedthrough

In high-frequency high-resolution applications the switch transistor is very large, even when bootstrapping is applied. The large transistor has large parasitic capacitances from the drain and the source to the gate and bulk nodes. In a bootstrapped switch the parasitic capacitances of the auxiliary circuitry have to be minimized, and thus no large devices are allowed. Consequently, the impedances shorting the gate and the bulk of the switch transistor to the ground during the switch off-phase are not negligible, which results in high-pass feedthrough paths past the closed switch, as illustrated in Figure 6.23 (a). Feedthrough is a severe problem, especially in S/H circuits, where the high frequency signal is present in the input whether the switch is
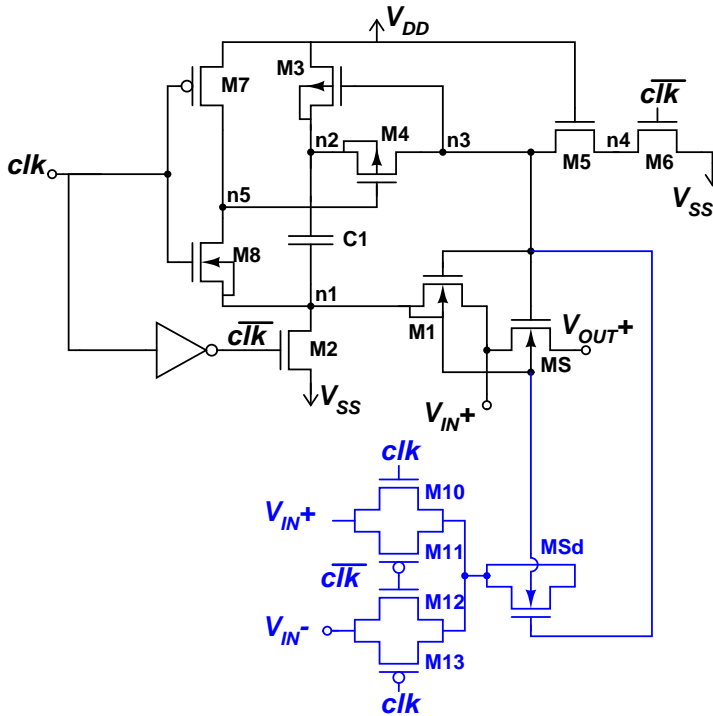
Figure 6.24 A bootstrapped switch with feedthrough cancellation.

open or closed.

A technique for reducing feedthrough is proposed in Figure 6.23 (b). It is based on the idea that in differential circuits the complementary signal can be used to cancel the feedthrough by generating a signal with equal magnitude but opposite phase. This is done by forming capacitors equal to C1 and C3 with a dummy transistor half the size of the switch and connecting these capacitors to the complementary input signal. The cancellation achieved is remarkable, but not perfect because of the signal dependency of the capacitor values.

In Figure 6.24 the proposed technique is applied in a bootstrapped switch realized with a triple-well process. During the switch on-phase the dummy transistor is connected to the same input as the switch to make the voltage over the junction diodes constant, which is essential in order to avoid distortion. Besides feedthrough reduction, the technique also brings another advantage: in the gate voltage equation (6.17) the term proportional to the input voltage at the end of the off-phase is also canceled. This switch, as a double-side bootstrapped version, has been employed as an input switch in the IF-sampling ADC presented in Section 12.5.

6.4.3.6   Bootstrapped Switch as a Sampling Switch

Thus far the presented circuits are intended to be used as an input switch (switch S_1 in Figure 6.10) in a circuit using the bottom plate sampling technique. Since the sampling switch (S_2) is connected to the ground from its other end, it does not see the signal swing and thus its on-resistance is allowed to be signal-dependent. In high frequency applications, however, the signal current may have such a large amplitude that it produces a significant voltage drop in sampling switch on-resistance (For example, a 100-MHz, 1-Vpp signal sampled in a 5-pF capacitor with a switch which has 10-Ω on-resistance results in a 31-mVpp voltage drop). As a result, the charge released when the switch is opened is not totally signal-independent. Bootstrapping the sampling switch reduces the on-resistance as well as makes the charge injection less signal-dependent.

To be used as a sampling switch, the circuit shown in Figure 6.18 has to be slightly modified in order to guarantee that the charge escaping from the bootstrapping circuit does not distort the sampled signal charge. This can be done by turning off the transistor M1 slightly before the switch transistor MS, which can be realized by connecting M1's gate to *clk* instead of node n3. Now the switch M1 does not have an on-resistance as small and signal-independent as before but, because of the small signal swing seen by the sampling switch, this is not critical.

6.5   Sampling Function

In the previous discussion the effect of the finite on-resistance of a MOS switch was modeled with a limited bandwidth in tracking mode. The same thing can also be thought of in a different way. As a result of the low-pass filtering effect the voltage across the sampling capacitor depends not only on the instantaneous value of the input voltage but also on its previous values. When a sample is taken it is a weighted average of the input values from the beginning of the time to the sampling moment. In practice, only a short time period has to be taken into account. The weighting can be modeled with the sampling function and the averaging with integration, as done in equation (3.5).

In a real circuit the resistance of the MOS switch cannot be changed from its constant on-value (the nonlinear effects are neglected in this discussion) in tracking mode to infinite in zero time. This is due to the finite slope of the transistor gate voltage and the gate voltage feedthrough to the source and drain via the parasitic capacitances. Since there is no single time instant which can be said to be the sampling moment, the

sampling must be modeled with the sampling function.

A thorough analysis of the nMOS transistor sampling function based both on analytic formulation and simulations is presented in [146]. The authors define the aperture time as the time interval which covers 80% of the sampling function area. Their analysis shows that for very small sampling capacitor values the aperture time has an almost linear dependence on the gate voltage fall time. On the other hand, when the sampling capacitor is large, the sampling function is dominated by the tracking mode time constant, which is typically the case in S/H circuits. Sub-samplers and digital high-speed line receivers are applications where the bandwidth reduction resulting from the finite turn-off time is remarkable.

Chapter 7

Operational Amplifiers

The opamp is a widely used building block in many types of analog circuits. Often, it is just the spot where the limits of the technology are first met, when trying to enhance the speed or reduce the power consumption of the circuits. In the SC technique and in the pipelined ADC based on it the opamp is a central component.

Opamp design methodology and various circuit topologies have been thoroughly covered in many textbooks. Thus, presenting a comprehensive study in this context does not serve the present purpose; indeed it may not even be possible. Instead, this chapter tries to concentrate on issues related to low-voltage and high-speed design with modern IC technologies. The requirements of opamps in SC circuits are reviewed and the most important and suitable circuit topologies, with their pros and cons, are compared.

7.1 Requirements for SC Applications

The maximum speed and, to a large extent, the power consumption of SC circuits are determined by the opamp. The opamps in SC circuits have some unique requirements, the most important of which is the input impedance, which must be purely capacitive so as to guarantee the conservation of charge. Consequently, the opamp input has to be a MOSFET, either in the common source or the source follower configuration. Since it is not possible to employ a BJT as an input transistor, the speed and power savings offered by the BiCMOS technology cannot be fully exploited.

7.1.1 Output Impedance

Another characteristic feature of SC circuits is the load at the opamp output, which is typically purely capacitive. As a result, since there is no need to drive resistive loads, the opamp output impedance can be high, making it possible to use operational transconductance amplifiers (OTAs). The output stage of an OTA provides significant voltage gain, enabling the achieving of the gain target with a smaller number of stages.

7.1.2 Output Voltage Range

The opamp output voltage range, as already discussed in Chapter 2, has a major impact on the signal-to-noise ratio. Consequently, maximizing the voltage swing is especially important in low-voltage and high resolution applications. Unfortunately, an output stage with high signal swing cannot usually provide very high output impedance, thus increasing the number of opamp stages. High output swing may also increase noise, since the output stage current sources cannot be optimally sized for low noise.

In fully differential opamps the common mode voltage level at the output is not automatically determined. To set it to the wanted level (typically in the middle of the supply voltages) a common mode feedback (CMFB) circuit has to be used.

7.1.3 Input Common Mode Range

SC circuits typically employ opamps in the inverting feedback configuration (and are fully differential, making signal inversion possible simply by crossing the wires), which does not require a large common mode voltage range in the opamp input. Thus, low-voltage circuits can be constructed with opamps that do not have a rail-to-rail input stage. If present, the single-ended to differential conversion in the circuit front-end, however, may need an opamp in the non-inverting configuration. In addition, a fully differential input signal without an exactly-known common mode level requires some common mode input range from the opamp of the front-end stage to accommodate signal CM voltage uncertainty or changes in it.

In SC circuits the opamp input common mode level need not be equal to the output CM level, which is usually set to $V_{DD}/2$ so as to maximize signal swing. This freedom can be utilized in low voltage circuits, for example when an nMOS input pair is employed, by setting the CM level close to the V_{DD}, which leaves more voltage headroom for the input pair and the tail current source. In principle, the CM level can be raised all the way to V_{DD}, but then care must be taken that the junction diodes of the pMOS switches attached to the opamp input do not become forward biased during

the settling.

7.1.4   DC Gain

The ultimate settling accuracy is limited by the finite opamp DC gain. What the exact settling error is depends not only on the gain but also on the feedback factor in the circuit utilizing the opamp. Typically, the DC gain requirement is from 60 dB up to 100 dB. In some circuits, such as a front-end S/H circuit, insufficient opamp DC gain results only in a gain error which is usually tolerable. The DC gain, however, has to be constant over the opamp output voltage range in order to avoid harmonic distortion.

7.1.5   Bandwidth and Phase Margin

When using a single pole model for the opamp, the settling time is determined by the gain-bandwidth product (GBW) of the opamp and the feedback factor of the circuit. In practical circuits, there is almost always more than just one pole and often zeros as well. However, in order to use the opamp in a closed loop configuration, it has to be designed in such a way that its frequency response is close to the single pole response. Consequently, there is one dominant low-frequency pole, while the other poles and zeros lie at much higher frequencies. In the frequency response their presence is seen as a phase roll-off in the high frequencies. Thus, the phase margin at the unity gain frequency has an effect on the settling time as well.

If the opamp is not utilized in unity gain feedback (e.g. auto-zeroing) the required phase margin is not defined at the unity gain frequency but at the frequency of the closed loop gain, and so it is easier to achieve.

The fastest settling is obtained when the first overshoot of the step response just touches the upper settling bound [147]. The higher the accuracy requirement, the closer the optimum becomes to the critically damped settling which corresponds to a 76° phase margin in a two-pole system. This is significantly larger than the 60° rule of thumb used for continuous time circuits and mentioned in some textbooks in conjunction with SC circuits.

Sometimes, when there is more than one non-dominant pole, zeros, pole-zero doublets, or complex pole pairs in the circuit, the phase margin does not give a good indication of the settling time; it can be significantly longer or shorter than in a two-pole system with the same phase margin.

7.1.6 Slew Rate

Besides the opamp bandwidth, the settling time is limited by the fact that the opamp can supply only a finite current to the load capacitor (or the compensation capacitor). Consequently, the output cannot change faster than the slew rate, which is given by

$$SR = \frac{I_{SR}}{C_L},\qquad(7.1)$$

where C_L is the load capacitance and I_{SR} the available slewing current. When designing an opamp, the load capacitor is known and the required slew rate ($SR = k \cdot V_{max}/T_S$) can be calculated from the largest voltage step (V_{max}) and the clock period (T_S), half of which is typically available for settling. A commonly used rule of thumb suggests that one third of the settling time should be reserved for slewing, resulting in k of six.

The required slewing current is

$$I_{SR} = \frac{k \cdot V_{max}C_L}{T_S}.\qquad(7.2)$$

It is linearly dependent on the clock frequency, while the current needed to obtain the opamp bandwidth has a quadratic dependence, which means that in high speed circuits the opamp current often needs to be higher than required by the slew rate [4]. On the other hand, in slow- and medium-speed circuits the slew rate keeps the current unnecessarily high. Since the slewing current is needed during only a fraction of the clock period, remarkable power savings can be achieved if the current can be adjusted according to the need. This can be accomplished either by using a class AB output stage or by dynamic biasing. The latter, however, has not gained popularity, while the class AB output stages are widely employed.

7.1.7 Noise

In high speed Nyquist-rate ADCs opamp noise is dominated by thermal noise, while $1/f$ noise is less important. Consequently, for noise reasons, there is no point in using a pMOS input pair (which has inherently lower $1/f$ noise) in the opamp.

The total noise contribution of all the devices in the opamp is usually combined as a single voltage source at the amplifier input. Assuming the noise sources to be uncorrelated, the total noise is obtained as a root of the sum of the squares of the individual input-referred noise sources. The noise contribution of the devices in the opamp's first stage is the most significant, and usually the noise of the other stages can be neglected, since it is attenuated by the preceding voltage gain.

For MOSFETs the gate-referred thermal noise density is given by

$$\overline{v_n^2} = 4\gamma \frac{kT}{g_m}\Delta f, \tag{7.3}$$

where T is the absolute temperature, k Boltzmann's constant, Δf the differential frequency, g_m the transistor small signal transconductance, and γ the noise excess factor, which is 2/3 for long channel devices (L>1.7 μm). In short channel devices the hot carrier effects increase the noise, leading to a larger value of γ. It has been experimentally shown [148] that for 0.7-μm devices the value of γ ranges from 2.5 to 9, depending on the bias conditions. Analytical models for γ have been proposed in [149] and [150]— unfortunately, they are too complicated for hand calculations. In general, however, γ increases as the gate-source voltage decreases and/or as the drain-source voltage increases.

The input-referred noise of the opamp input pair is directly given by equation (7.3). Thus, it is reduced by increasing the transconductance, which can be done by utilizing nMOS devices, increasing the current, or increasing the aspect ratio of the devices. The effect of the last method, however, is partially canceled by the increase in γ.

When referred to the opamp input, the noise voltages of the transistors used as current sources (or mirrors) in the first stages are multiplied by the transconductance of the device itself and divided by the transconductance of the input transistor. As a result, the total input-referred noise is

$$\overline{v_{n,opamp}^2} = 4 \cdot \frac{kT\gamma_{in}}{g_{m,in}}\Delta f \cdot \left(1 + \frac{\gamma_{cs1} \cdot g_{m,cs1}}{\gamma_{in}g_{m,in}} + \frac{\gamma_{cs2} \cdot g_{m,cs2}}{\gamma_{in}g_{m,in}} + \cdots\right), \tag{7.4}$$

which again suggests that maximizing input pair transconductance minimizes noise. It can be further reduced by decreasing the transconductances of the current sources. Since the current is usually set by other requirements, the only possibility is to decrease the aspect ratio of the device. This leads to an increase in the gate overdrive voltage $V_{GS}-V_T$, which, as a positive side effect, also decreases γ. It should be noticed that the overdrive voltage is equal to V_{dsat}. Consequently, obtaining low noise with low supply voltage is difficult, especially with single stage opamps, where the output signal swing does not permit large V_{dsat}. Increasing L to avoid short channel effects is also possible, but with a constant aspect ratio it increases the parasitic capacitances, reducing the opamp bandwidth.

Cascode transistors do not make a significant contribution to noise, because their noise voltage is transformed into current through the high output impedance of the underlying current source.
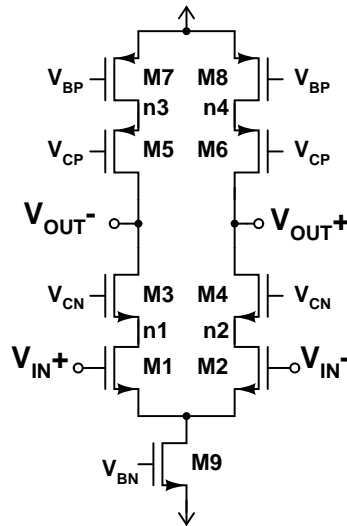
Figure 7.1 Telescopic OTA [120].

When looking at equation (7.4), it is clear that opamp topology does not affect the noise contribution of the input pair except through the device type (nMOS or pMOS). On the other hand, the number, and to some extent, the magnitude of the terms in the parentheses are dependent on the topology. When comparing opamp topologies it is convenient to write the term in parentheses as $1 + \gamma_{OA}$, where γ_{OA} is referred to as opamp noise excess factor.

7.2 OTAs with Single High-Gain Stage

OTAs with a single gain stage have been widely employed in SC circuits. A high output impedance provides an adequate DC gain, which can be further increased with gain boosting techniques. Single-stage architecture offers large bandwidth and a good phase margin with small power consumption. Furthermore, no frequency compensation is needed, since the architecture is self-compensated (the dominant pole is determined by the load capacitance), which makes the footprint on the silicon small. On the other hand, the high output impedance is obtained by sacrificing the output voltage swing, and the noise is rather high as a result of the number of noise-contributing devices and limited voltage head-room for current source biasing.

Figure 7.2 Folded cascode OTA [151].

7.2.1 Telescopic OTA

The telescopic OTA [120], shown in Figure 7.1, is probably the fastest possible architecture. Both the GBW and the lowest non-dominant pole are determined by nMOS devices, resulting in both large bandwidth and good phase margin. The number of current legs being only two, the power consumption is small.

The most prominent drawback of this architecture is the limited voltage swing both at the output and the input of the opamp. From the high side the output swing is limited to $2V_{dsat}$ below V_{DD} and from the low side a minimum of $3V_{dsat}$ above V_{SS}. With this maximum possible output swing the input common mode range is zero. In practice, some input CM range, which reduces the output swing, always has to be reserved so as to permit inaccuracy and settling transients in the signal common mode levels. With supply voltages of 5 V or larger the voltage swing, however, is often more than sufficient, so that even an extra set of cascodes can be inserted in both the nMOS and pMOS sides to enhance the DC gain. But, when the supply is 3 volts or less, the swing is too small for most SC applications.

7.2.2 Folded Cascode OTA

The folded cascode OTA [151], shown in Figure 7.2, is probably the most commonly used opamp architecture in SC circuits. It provides a larger output swing and input CM range than the telescopic OTA with the same DC gain and without major loss of speed. The output swing, $V_{DD} - 4V_{dsat}$, is not linked to the input CM range, which is $V_{DD} - V_T - 2V_{dsat}$ (obtained using $V_{GS} = V_T + V_{dsat}$).

The choice between an nMOS and pMOS input pair has to be made on the basis of the required phase margin. The nMOS input architecture, shown in Figure 7.2, offers large GBW (g_{m1}/C_L) thanks to the nMOS input transistors, but the lowest non-dominant pole (g_{m6}/C_1) associated with the node n1 is determined by the low pMOS transconductance and the large stray capacitances of the pMOS current sources and the cascode devices. On the other hand, utilizing a pMOS input pair gives lower GBW, but the non-dominant pole is higher, thanks to the nMOS cascode devices.

Feedforward capacitors can be used to bypass the cascode transistors at high frequencies to improve the phase margin [152, 153, 154, 155]. In principle, the technique produces a zero, which is used to cancel the pole associated with the cascode node. It is, however, not possible to place this zero exactly on top of the pole. Thus, there is a sufficiently closely spaced pole-zero pair, a doublet, which is known to introduce a slowly settling component in the step response [156]. Consequently, feedforward techniques should be used with care in opamps employed in SC circuits.

It is possible to employ an nMOS and a pMOS input pair in parallel [157], which increases the slew rate by 1/3 (with the same total current consumption), but at the same time increases the input capacitance and thermal noise and lowers the non-dominant pole. Another possible way to increase the slew rate and ensure that all transistors remain in saturation during slewing is to clamp the cascode nodes with diode-connected devices [158].

An important fact, which is not mentioned in many textbooks, is that impedance seen at the cascode node is actually high at the DC [159]. The impedance is equal to the output impedance of the opamp, attenuated by the gain of the cascode device (g_m/g_{ds}). Thus, in a folded cascode structure the impedance of the cascode node is in the order of ($2/g_{ds}$). When going to higher frequencies the load capacitance starts to dominate the opamp output impedance and, as a result, the impedance at the cascode node decreases without forming a low-frequency pole in the frequency response.

In SC circuits high cascode impedance can be harmful, since it results in a Miller multiplication of the gate-drain capacitance of the opamp input device. In an SC amplifier a significant amount of charge intended for the feedback capacitor ends up in the opamp input capacitances, resulting in a gain error. The Miller effect can be avoided by inserting extra cascode transistors on top of the input pair [160, 161]. In the case of an nMOS input pair, the added non-dominant pole is much higher than the one already present in the transfer function and thus the phase margin is not reduced significantly. An alternative method to get rid of the Miller effect is to put capacitors, matched to the gate-drain capacitance of the input device, between the gate of the input transistors and the drain of the complementary input transistor [70].
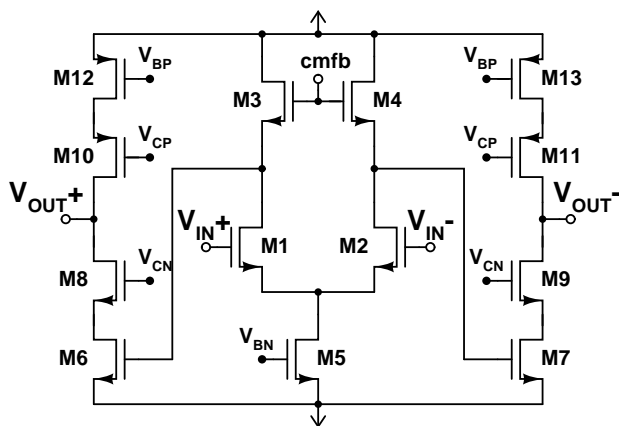
Figure 7.3 OTA with low-gain preamplifier [73].

In BiCMOS technology the cascode transistors in a folded cascode opamp can be implemented with bipolar devices, resulting in a considerably high non-dominant pole.

7.2.3 Cascode Stage with Low-Gain Preamplifier

In addition to folding, another way to get the input pair current to the cascode output stage is current mirroring. The resulted circuit provides a better slew rate than the folded cascode, but introduces another non-dominant pole, which becomes lower as the current mirroring ratio is increased.

A closely-related architecture is shown in Figure 7.3 [73]. Instead of a current mirror, the first stage load is a pair of common gate-connected nMOS devices (M3 and M4) and the signal is taken into the output stage from the nMOS side. As a result, the use of pMOS devices in the signal path is avoided, giving a large GBW and pushing the non-dominant poles into high frequencies. There are, however, two non-dominant poles, one associated with the first stage output and the other with the cascode node, which makes the phase roll-off steep once it begins. To make sure that the first stage output pole is high enough, the g_m of M3 and M4 has to be large, limiting the first stage gain (or current mirroring ratio) to sufficiently small values, typically smaller than two. Unfortunately, the thermal noise increases as the first stage gain is decreased and thus this topology is unsuitable when low noise is required. Also, the input CM range is fairly limited.

In BiCMOS technology the output stage nMOS devices can be replaced with npn transistors, resulting in a very high GBW.

Table 7.1 Comparison of single-stage OTAs

	Telescopic OTA	Folded Cascode	OTA with LG preamp
DC gain	$\frac{g_m \cdot g_m}{g_{ds} \cdot g_{ds}}$	$\frac{g_m \cdot g_m}{g_{ds} \cdot g_{ds}}$	$\frac{g_{m1}}{g_{m3}} \cdot \frac{g_m \cdot g_m}{g_{ds} \cdot g_{ds}}$
GBW	$\frac{g_{m1}}{C_L}$	$\frac{g_{m1}}{C_L}$	$\frac{g_{m1}}{g_{m3}} \cdot \frac{g_{m6}}{C_L}$
2. pole	$\frac{gm3}{C_{n1}}$	$\frac{gm6}{C_{n1}}$	$\frac{gm3}{C_{n1}}, \frac{gm7}{C_{n3}}$
I_{SLEW}	I_S	I_S	mI_S
I_{VDD}	I_S	$2I_S$	$I_S + nI_S$
noise (γ_{OA})	$\frac{\gamma_7 g_{m7}}{\gamma_1 g_{m1}}$	$\frac{\gamma_4 g_{m4}}{\gamma_1 g_{m1}} + \frac{\gamma_{10} g_{m10}}{\gamma_1 g_{m1}}$	$\frac{g_{m3}}{\gamma_1 g_{m1}}\left(\gamma_3 + \frac{g_{m3}}{g_{m6}}\left(\gamma_6 + \frac{\gamma_{12} g_{m12}}{g_{m6}}\right)\right)$
out swing	$V_{DD} - 5V_{dsat} - v_{incm}$	$V_{DD} - 4V_{dsat}$	$V_{DD} - 4V_{dsat}$
in CM	0 to $V_{DD} - 5V_{dsat}$	$V_{DD} - V_T - 2V_{dsat}$	$V_T - V_{dsat}$
min V_{DD}	$V_T + 2V_{dsat}$ or $5V_{dsat}$	$V_T + 2V_{dsat}$	$2V_T + 2V_{dsat}$
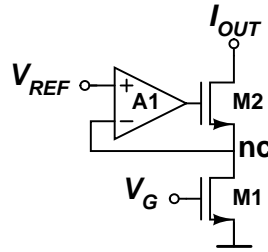
Figure 7.4 Regulated cascode current source.

7.2.4 Comparison of Single-Stage OTAs

The main characteristics of the OTAs presented above are collected in Table 7.1 using the following notations: g_{mx} is the transconductance of transistor Mx, C_L is the load capacitance, C_{nx} is the parasitic capacitance associated with node nx, and I_S is the input pair tail current. In the third OTA m is the ratio of the aspect ratios of M6 and M3 and n is the ratio of the output stage current to the input stage current. The noise is given with the noise excess factor γ_{OA}. It can be seen that each architecture has its pros and cons, suggesting that the choice has to be made on the basis of which specifications are important in the target application.

7.2.5 Gain Enhancement Techniques

In many applications the opamp DC gain requirement is higher than what is achievable with simple single-stage topologies. Techniques to enhance the opamp DC gain without going into multi-stage architectures are especially welcome in high speed circuits, where the high current levels make the transistor g_{ds} large.

A very widely-used method is based on improving the cascoding effect of a single
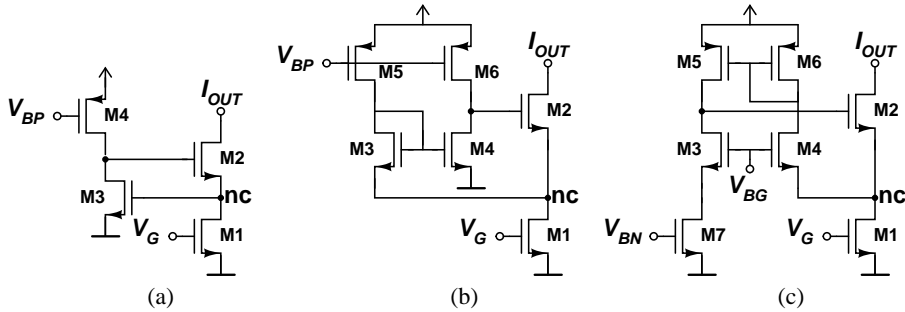
Figure 7.5 Regulated cascode current source: three implementations.

MOS transistor by using local negative feedback [162]. The resultant circuit, often referred to as regulated cascode, is utilized in a current source that is shown in Figure 7.4. There, the auxiliary amplifier encloses the cascode transistor M2 in a feedback loop, making the voltage on its source node almost constant. As a result, the output impedance of the current source is given by

$$r_{out} \approx \frac{A_1 g_{m2} g_{m1}}{g_{ds2} g_{ds1}}. \tag{7.5}$$

Thus, the regulation improves the impedance by the gain of the regulation amplifier A_1 and, when the current source is utilized in an OTA the DC gain is increased by the same amount.

Three different implementations of the regulation amplifier are shown in Figure 7.5. The one in Figure (a) [162, 163] is very simple, but sets the voltage on the cascode node unnecessarily high. The circuit in Figure (b) [164] utilizes a level shifter and the other one in Figure (c) [165] a common gate amplifier, to allow the biasing of the cascode node to a lower voltage. Using a more complicated regulation amplifier, e.g. a folded cascode OTA, is also possible. In fully differential circuits the regulation amplifier can also be fully differential.

The regulated cascodes were for the first time utilized in an opamp in [166], where the DC gain of a folded cascode OTA was boosted to 90 dB. It was shown that if the GBW of the regulation amplifier is larger than the dominant pole of the unregulated opamp, the regulation does not have a significant effect on the opamp bandwidth. However, it introduces a pole-zero doublet, which may slow down the settling to a remarkable extent. Thus, it was suggested that the regulation amplifier GBW should be larger than the closed loop bandwidth of the opamp.

The frequency response and the settling of the regulated cascode opamp were

analyzed in more detail in [167] with the aid of a symbolic circuit simulator. It was found that when the doublet is pushed into higher frequencies it in fact merges with the non-dominant pole of the opamp, forming a complex pole pair and a zero. Achieving fast settling with a reasonable cost requires the optimizing of the pole pair quality factor, frequency, and the zero frequency with respect to each other. Optimal settling calls for a regulation amplifier GBW, which is substantially higher than the opamp GBW. An opamp designed according to these principles has been utilized in [168].

When the regulation amplifier is more complex than the one in Figure 7.5 (a) it has non-dominant poles, which make the frequency response and the settling behavior more complicated to analyze.

In addition to the cascode regulation other techniques for increasing the DC gain have been proposed as well. Gain boosting with positive feedback has been investigated, e.g. in [169] and [170]. In [171], dynamic biasing, where the opamp current is decreased toward the end of the settling phase, is used to increase the DC gain. It exploits the fact that current reduction lowers the transistor g_{ds}, which increases the DC gain.

In a feedback configuration, due to finite opamp gain, the opamp input is not a perfect virtual ground but sits on the voltage $-V_{OUT}/A_0$. Reducing this signal-dependent value by some means has the same effect as increasing opamp DC gain.

In [173] a replica circuit with an opamp plus feedback circuit is used to generate the voltage difference in the input terminals. A second replica opamp, which has its output tied to the output of the main opamp, uses the generated voltage as its input to drive the load. The main amplifier only needs to fine-tune the output voltage and as a result, its effective DC gain is enhanced by $A_0/(1+\beta)$, where β is the feedback factor. This technique is also suitable for circuits with a resistive load.

In SC circuits the same principle can be realized in the time domain by having an extra clock phase prior to the amplification phase to sample the input voltage in a capacitor put in series with the opamp input [172, 94].

7.3   Two-Stage Opamps

In high-resolution low-voltage applications thermal noise and the opamp output swing are emphasized. Thus, two-stage opamps are often preferable to single-stage ones.
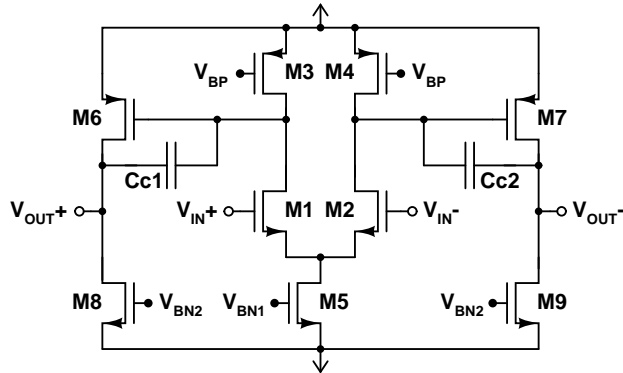
Figure 7.6 Miller opamp.

7.3.1 Miller Opamp

The Miller opamp is the most basic two-stage opamp. It provides rail-to-rail output swing and low thermal noise and can be used with supply voltages down to $V_T + 3V_{dsat}$. The GBW is g_{m1}/Cc and the non-dominant pole g_{m6}/C_L, which is lower than in single-stage OTAs. Consequently, when a large phase margin is required, the Miller opamp cannot achieve as a large bandwidth as single-stage OTAs.

Often, a resistor is put in series with the compensation capacitor in order to get rid of the right half plane (RHP) zero that results from the feedforward path through the capacitor.

7.3.2 High-Gain First Stage and Rail-to-Rail Output Stage

The basic Miller topology offers only a moderate DC gain, which is often too small for high resolution applications. The most straightforward way to increase the gain is to employ a high-gain first stage. Since the first stage does not need to have a large output voltage swing, it can be a cascode stage, either a telescopic or a folded cascode. The advantages of the folded cascode structure (Figure 7.7) are a larger input CM range and the avoidance of level shifting between the stages, while the telescopic stage (Figure 7.8) can offer larger bandwidth and lower thermal noise. Thus, when there is no special need for large input CM range the latter is more attractive.

Level shifting between the telescopic input stage and the rail-to-rail output stage can be avoided by taking the signal into the output stage from the pMOS side. However, it has two disadvantages. First, the voltage across the first stage cascode current sources would be only $V_T + V_{dsat,12}$, which does not permit the optimal biasing of the current sources for low noise. Second, the low transconductance of the pMOS device

Figure 7.7 Folded cascode OTA with rail-to-rail output stage.

Figure 7.8 Telescopic OTA with rail-to-rail output stage.
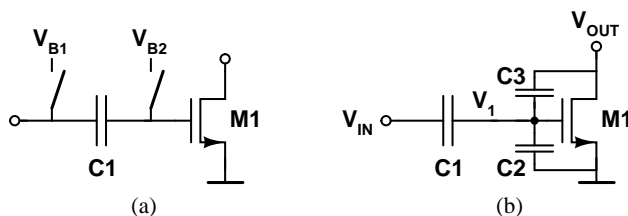
Figure 7.9 Capacitive level shifting: (a) principle and (b) with parasitic capacitances.

results in a low-frequency non-dominant pole.

The level shifting can be realized with a source follower. It will, though, increase current consumption and noise. In SC circuits a capacitive level shifter, which does not have these disadvantages, can be utilized [174]. Figure 7.9 (a) shows the principle of a switched capacitor level shifter, which is based on a capacitor that is periodically charged to the desired offset voltage ($V_{B1} - V_{B2}$).

The parasitic capacitances, shown in Figure 7.9 (b), form a capacitive voltage divider with the level shift capacitor C1. Since, as a result of the Miller effect, the gate-drain capacitor C3 is multiplied by the gain of the output stage, the capacitor C1 has to be large to avoid attenuation. It can easily be shown that with capacitive level shifting the output stage DC voltage gain is

$$\frac{V_{OUT}}{V_{IN}} = \frac{g_{m1}C_1}{g_o\left(C_1 + C_2 + C_3 + \frac{g_{m1}}{g_o}C_3\right)}, \tag{7.6}$$

which is always less than C_1/C_3. With a sufficiently large C1 this is not a severe limitation.

7.3.2.1 Frequency Compensation

When there are cascode nodes in the first stage, cascode compensation, in which the compensation capacitor is connected to the cascode node instead of the first stage output, can be used instead of the standard Miller compensation [175, 176]. It offers higher GBW and the RHP zero is at a much higher frequency.

Instead of a single non-dominant pole there is a complex pole pair and a zero [176]. It is well-known that, because of a complex pole pair, cascode compensation can result gain peaking near the unity gain frequency, increasing the settling time. To avoid this, the Q-value of the complex pole pair has to be less than $1/\sqrt{2}$, which is

satisfied when [177]

$$g_{mc} > \frac{2 \cdot C_c \cdot m}{C_2 \left(1+m\right)^2} \cdot g_{m2,} \tag{7.7}$$

where g_{mc} is the transconductance of the cascode device, g_{m2} the transconductance of the output stage input device, C_c the compensation capacitor, C_2 the capacitance from the first stage output to the ground, and m the ratio of the load capacitor to the compensation capacitor. When the load capacitor is equal to the compensation capacitor the equation reduces to $g_{mc} > 0.5 \cdot g_{m2} \cdot C_C/C_2$. Typically, when the capacitor ratio C_C/C_2 is of the order of ten or more, this is not easy to achieve. The situation is somewhat easier if the cascode device has inherently larger g_m than the output device (e.g. nMOS vs pMOS or BJT vs MOSFET).

In [177] it is shown that a good compromise between bandwidth and gain peaking can be achieved by combining the standard Miller and cascode compensation. In [178] simulations show that using two compensation capacitors, one connected to the cascode node in the signal path and the other to the cascode node in the current source of a telescopic first stage, yields faster settling than a single capacitor connected to either node alone.

7.3.2.2 Two-Stage BiCMOS Opamp

A two-stage BiCMOS opamp designed to achieve high speed, high DC gain, and high SNR is shown in Figure 7.10. This opamp is utilized in the ADC prototype [1] described in Section 12.5.

The telescopic input stage provides high DC gain and minimizes the number of noise-contributing devices. The capacitive level shifting between the stages avoids the use of pMOS devices in the signal path and allows for the optimal biasing of the first stage common mode level for high DC gain and low noise. The level shift capacitors are charged during one half of the clock period by switching one capacitor terminal to a bias voltage, generated with a dummy structure, and letting the first stage common-mode feedback set the voltage at the other terminal. The cascode devices are implemented with bipolar transistors with an inherently much larger g_m than the MOS transistors. Consequently, the cascode compensation can be used without the need to be concerned about gain peaking. The input stage and the output stage have separate common mode feedback loops for two reasons: first, it is easier to achieve fast and accurate CMFB without a fear of instability with separate loops and, second, the first-stage CMFB is used to charge the level shift capacitor, which is not possible with a single loop since the signal path between the stages is broken during charging. Both the CMFB circuits are realized with standard SC circuits [179].
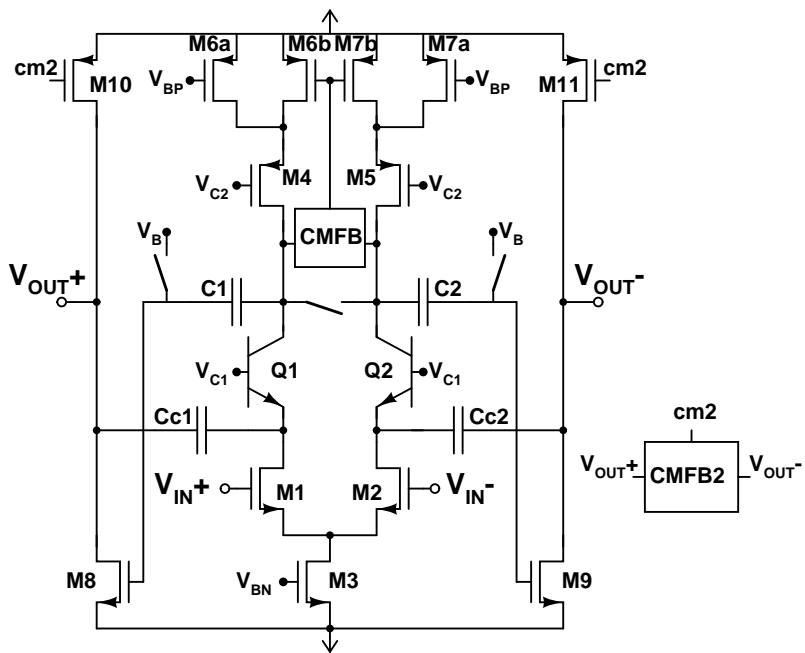
Figure 7.10 High-performance BiCMOS opamp.

Chapter 8

Clock Generation

The operation of different circuit blocks in SC circuits is synchronized with a clock signal. To guarantee that the signal at the output of an opamp is sampled in a steady state, non-overlapping clock signals are used in adjacent stages. As long as the sampling clock edge arrives before the one that ends the hold-phase in the driving stage, the circuit is very robust against small timing uncertainty at the clock edges. In the front-end, where the clock is used to sample a continuous time signal, the situation is completely different.

Any deviation of the sampling moment from its ideal value results in an error voltage in the sampled signal. The error is equal to the signal change between these two moments. Thus, the sampling clock has to be regarded as a sensitive analog signal and treated accordingly.

The sampling moment is determined by the clock signal zero crossing, which corresponds to the moments when the clock phase is an integer multiple of 2π. Random variations in the phase, also known as phase-noise, are a source of timing errors. The phase-noise is usually specified with the single-sideband noise spectral density $\mathcal{L}(f)$, which, being a frequency domain parameter, does not always give a good insight into the timing error. A related time domain parameter, jitter, is defined as an rms error between a reference time point and the clock signal zero crossing. Cycle-to-cycle jitter is often of interest and the previous zero crossing is taken as the reference point.

Besides random errors, the signals present on the chip or the circuit board can couple to the clock.

8.1 Jitter

The instantaneous voltage error caused by jitter Δt can be approximated as

$$V_n \approx \Delta t \cdot \frac{dV_S(t)}{dt} = \Delta t \cdot A_S 2\pi f \cos(2\pi f t), \qquad (8.1)$$

where $V_S(t)$ is the signal waveform, which in the second form is assumed to be a sinusoid with amplitude A_S and frequency f. The rms voltage error can be obtained by squaring (8.1) and integrating it over the signal period, yielding

$$\overline{V_n} = \Delta t \cdot A_S \sqrt{2}\pi f. \qquad (8.2)$$

Using this, the signal-to-noise ratio can be calculated to be

$$SNR = -20 \cdot \log\left(2\pi f \Delta t\right). \qquad (8.3)$$

It can be seen that the *SNR* is independent of signal amplitude and that it decreases as the signal frequency increases. The maximum allowed jitter for a certain target *SNR* is given by

$$\Delta t \approx \frac{1}{f \cdot 10^{\frac{SNR+16}{20}}}. \qquad (8.4)$$

Knowing that the quantization noise power in an ADC is $LSB^2/12$ and allowing the error power generated by jitter be equally large (in the case of full-scale sine wave input) the jitter requirement can be written as

$$\Delta t \approx \frac{1}{f \cdot 7.7 \cdot 2^N}. \qquad (8.5)$$

If only 1 dB SNR degradation is permitted, the factor 7.7 increases to 15.1.

8.1.1 Jitter Sources

The clock signal is usually taken from a crystal oscillator or from the output of a phase locked loop (PLL), locked to a crystal reference. A typical oscillator phase noise spectrum is shown in Figure 8.1 [180]. From it, three different regions can be identified; in the vicinity of the carrier (small Δf) the noise power is proportional to $1/\Delta f^3$. When going further in frequency the slope changes to $1/\Delta f^2$ and eventually the noise spectrum becomes flat. The noise power in the non-flat region is inversely proportional to the oscillator Q-value and power dissipation. The crystals have very
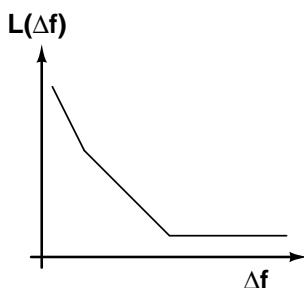
Figure 8.1 Typical oscillator phase noise spectrum.

large Q's and thus small phase noise and jitter.

In a PLL the phase noise of the voltage-controlled oscillator (VCO) is attenuated by the loop filter; the wider the filter bandwidth, the faster the loop correcting action and the resultant attenuation. On the other hand, the effect on the reference phase noise is just the opposite; smaller bandwidth gives more attenuation. Typically, however, the VCO phase noise dominates.

When sampling a sinusoidal input signal with a clock signal whose noise spectrum has the shape shown in Figure 8.1, the resultant signal spectrum will have a similar shape. Thus, in general, close-in-carrier phase noise spreads the signal spectrum, while white wide-band phase noise raises the noise floor of the sampled signal. Many applications set distinct requirements for them. For example, in a narrow channel communication system the spreading of a strong interferer because of phase noise may mask a nearby weak channel, leading to strict specifications for close-in phase noise.

The jitter, being a single number, does not contain as much information as the phase noise and thus is not always a sufficient figure of merit on its own. The thermal noise originating after the oscillator, e.g. in the buffering, does not get accumulated in the phase and thus the resulting phase noise has a white spectrum. Consequently, it can be fully described by the jitter. Similarly, the jitter of a crystal oscillator is also typically dominated by white phase noise.

There are not many ways to reduce the effects of jitter. Oversampling can be used to spread the white noise arising from jitter over a wider frequency range, which allows the *SNR* to be improved with discrete time (analog [181] or digital) filtering and decimation. Increasing the oscillator frequency also tends to reduce the absolute jitter, since for a constant single-side-band noise, the jitter is inversely proportional to the oscillator frequency. In addition, it is important to avoid unnecessary buffering of the sampling clock.
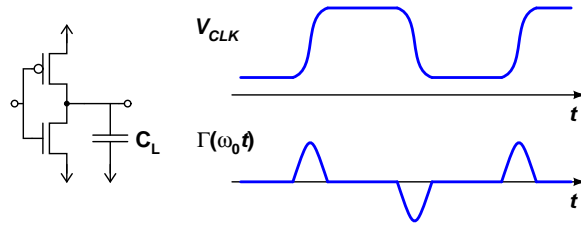
Figure 8.2 Clock waveform and corresponding impulse sensitivity function (ISF).

8.1.2   Inverter Buffer

On-chip buffering for the clock signal is needed primarily for two reasons. First, the capacitive loading in the clock lines may be high, requiring a buffer to keep the clock edges sharp. Second, the incoming clock cannot usually be directly utilized, but it is used as an input for a clock generator, which produces the non-overlapping clock signals. After the clock generator, the signals are buffered to achieve the required driving capacity. The buffering is the critical place where a clock signal, which originally has a low jitter can easily be contaminated.

The simplest buffer is the CMOS inverter. How its voltage noise turns into jitter is investigated next. It is clear that with a large amplitude input signal the inverter is a highly nonlinear circuit, and thus it has to be carefully considered, whether the linear noise models are applicable. In conjunction with the ring oscillators [182], the transformation of thermal noise into jitter is modeled with a linear time-varying model. It is based on the observation that the sensitivity of the zero crossing moment to a noise impulse varies over time (or phase), being largest when the inverter is changing its stage and becoming practically zero when the output voltage is saturated to the high or low logic level. The phenomenon is modeled with a time-varying dimensionless impulse sensitivity function (ISF) Γ, which is illustrated in Figure 8.2.

Here, the analysis is simplified by assuming that the inverter output slews during the whole transient and that the impulse sensitivity function is a square pulse with the same length as the transient. Consequently, a linear time-invariant model is used.

The thermal noise is modeled with a current source connected to the output and having the value $\overline{i_n} = \sqrt{4kT\gamma g_m \Delta f}$. The load of the inverter is a parallel combination of the load capacitance and the output resistance of the inverter: $Z_L = (1/r_o + sC_L)^{-1}$. The noise current is transformed to voltage in the output impedance, resulting in

$$\overline{V_n} = \sqrt{\frac{kT\gamma g_m r_o}{C_L}}, \tag{8.6}$$

which is obtained by integrating the spectral density from the zero frequency to infinity. The amount of time which the zero crossing moves as a result of the noise voltage is inversely proportional to the slope of the signal. Since the inverter is slewing, the slope is equal to the slew rate, resulting in the following equation for the jitter:

$$\Delta t = \frac{\overline{V_n}}{SR} = \frac{\sqrt{kT\gamma g_m r_o C_L}}{I_D} = \sqrt{\frac{2kT\gamma r_o C_L}{\frac{W}{L}\mu C_{ox}(V_{DD}-V_T)^3}}, \tag{8.7}$$

where $SR \, (= I_D/C_L)$ is the slew rate. The last form of the equation is obtained using the square law model for the transistor current. In [183] an analysis of a differential delay cell yielded similar results.

Probably the main finding in this jitter equation is its dependence on inverter DC gain ($g_m r_o$), which is due to the fact that at low frequencies the noise current is transformed into voltage in the output resistance. Since the signal slope is determined by the slew rate, not the gain, there is no reason for having a large output resistance, which suggests that even lowering it with an additional parallel resistor might be useful. Another, fairly unsurprising finding is the fact that increasing the current reduces the jitter. Furthermore, the last form of the equation shows clearly how advantageous it is to have a large supply voltage.

8.2 Signal Crosstalk

One strong interfering signal, typically present both at the board and the chip level, is the analog input signal itself. Capacitive or inductive coupling between it and the clock can happen between the traces on the PCB, the package pins, or the bonding wires. On the chip, coupling through the substrate or via modulation of the supply voltage is possible. Fully differential circuitry reduces both the coupling and the sensitivity to it and thus it should be used for both the input signal and the clock, if possible.

If the clock signal couples to the input signal, the sampling will alias the fundamental clock frequency and all its harmonics to the DC. The resulting DC offset is generally not harmful.

On the other hand, if the input signal couples to the clock, the sampling produces a spurious signal at twice the signal frequency, which will be shown next.

Let the ideal clock waveform be $V_{CLK,id}$ and the signal V_{IN}. When the signal couples to the clock circuitry the resulting sampling clock is $V_{CLK,id} + B(f) \cdot V_{IN}$, where $B(f)$ is a frequency-dependent coupling factor. The situation in the vicinity of the clock waveform zero crossing is illustrated in Figure 8.3. Assuming that the ideal
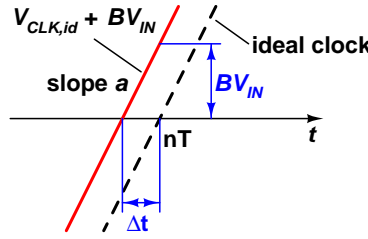
Figure 8.3 Input signal coupling to the clock signal and its effect on the sampling moment.

clock signal is linear with slope a, the signal after the coupling is:

$$V_{CLK} = a(t - nT) + B(f) \cdot V_{IN}(t), \tag{8.8}$$

where nT is the ideal zero crossing moment. Solving t and assuming that the signal does not change much between the ideal and the actual zero crossings yields

$$t \approx nT - \frac{B(f) \cdot V_{IN}(nT)}{a} = nT - \Delta t. \tag{8.9}$$

It can be seen that the situation is analogous to the one, studied in Section 6.1 in conjunction with signal-dependent switch turn-off moment. Using the result derived there, the relative level of the spur at two times the signal frequency can be written as

$$HD_2 = -20 \cdot \log \left(\frac{B(f) A_S f_S \pi}{a} \right), \tag{8.10}$$

where A_S is the amplitude and f_S the frequency of the input signal. It can be seen that level of the spur rises with signal frequency and amplitude, and it can be reduced by reducing the coupling and by making the clock waveform steeper. It should be noted that the clock slope of interest is where the coupling, not the sampling, happens. In addition, since the spur is rather a result of mixing than of amplitude distortion, fully differential circuitry does not help once the clock has been contaminated.

8.3 Circuits

8.3.1 Standard Non-overlapping Clock Generator

The clock generator for producing non-overlapping clock signals can be realized with a simple circuit constructed of logic gates. Such a circuit is shown in Figure 8.4. It is based on the idea that the falling edge of the input clock passes immediately
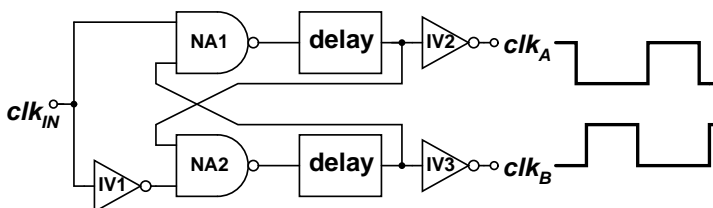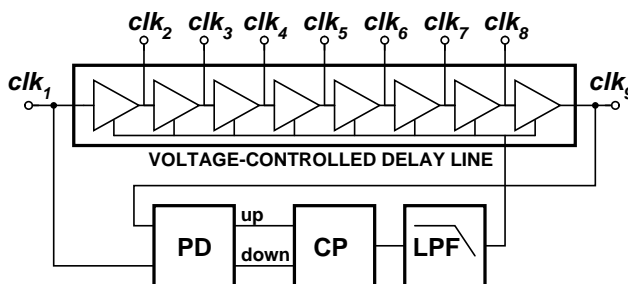
Figure 8.4 A simple clock generator.

Figure 8.5 Delay locked loop (DLL).

through the NAND gate NA1, while the rising edge has first to propagate through the other NAND gate and the cascaded delay element. The resulted non-overlapping signals clk_A and clk_B have a non-overlapping time equal to the sum of the delays at the NAND gate and the delay element. The delay element is usually realized with an even-numbered chain of inverters.

The main advantage of this circuit is its simplicity. At least a part of the buffering of output signals can be included in the delay elements, making the circuit quite robust. On the other hand, the non-overlap time often becomes larger than necessary because of the buffering included and the margin added to accommodate the process and temperature variations. The resulting speed penalty is emphasized in high clock rate circuits. Furthermore, the duty cycle of the generated clock signals is inherited from the input clock, requiring it to be close to 50%.

8.3.2   DLL-Based Clock Generator

A process and temperature independent non-overlap time can be realized using a delay locked loop (DLL). Figure 8.5 shows a simplified block diagram of a DLL. It consists of a voltage-controlled delay line, a phase detector (PD), a charge pump (CP), and a loop filter. Negative feedback is utilized to adjust the variable delay to be equal to the clock period (or its half).
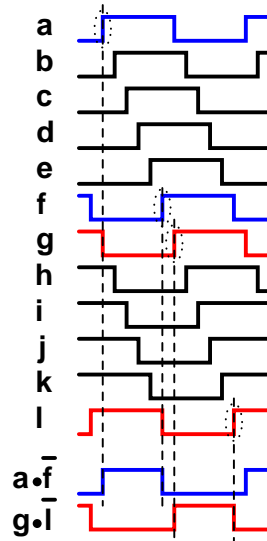
Figure 8.6 Generation of non-overlapping clock signals using edge combining.

In the feedback loop the phase detector generates *up* and *down* pulses, which are proportional to the phase difference between the input and the output clock. The charge pump performs a D/A conversion on the pulses and the low pass filter averages them over some time. The filter output voltage controls the delay line, eventually forcing the phase difference to zero.

The delay line consists of several unit elements, which divide the total delay into uniformly spaced sub-intervals, the number of which is equal to the number of elements. The output signals of the individual elements are available outside the delay line. Figure 8.6 shows how these signals can be utilized to form non-overlapping clock signals. The rising edges of four signals are combined with the logic AND operation, resulting in two new signals, whose non-overlap time is equal to the unit delay and hence a fixed portion of the clock period.

The DLL-based clock generator is clearly a more complex circuit than the circuit based on NAND gates. Consequently, it consumes more power and area and requires greater design effort. Thus, its benefits have to be carefully weighed against the disadvantages when the design decision is being made.

The utilization of a DLL can be more versatile than just constructing non-overlapping signals. For example, in the time-interleaved pipelined ADC presented in [93] the need for a full-speed clock signal is avoided by utilizing a DLL to generate all the required clock phases for the parallel channels. A DLL can be used to generate

signals with virtually any duty cycle and even as a frequency multiplier.

When very low jitter is essential, it is advantageous not to put the sampling clock through a clock generator, but take it as early after it arrives on the chip as possible. A DLL can be employed to construct the required complementary non-overlapping signal, as is done in the ADC prototype described in Section 12.5.

Chapter 9

Double-Sampling

The clock rate of the switched capacitor circuits is limited by the bandwidth of the opamp; thus, in order to achieve a high speed it is essential to exploit the opamp efficiently. This chapter introduces a technique to double the sampling rate of the switched capacitor circuits without a need to increase the speed of the opamp. This technique, called double-sampling, was first introduced in [184]. It has been applied in various SC circuits such as filters, $\Delta\Sigma$-modulators [185, 186], pipelined ADCs [97, 4], and S/H circuits [11, 10, 9, 187].

This chapter begins with the introduction of double-sampling technique, after which its nonidealities are analyzed. A circuit structure for eliminating one of those, the timing skew, is proposed in the last section.
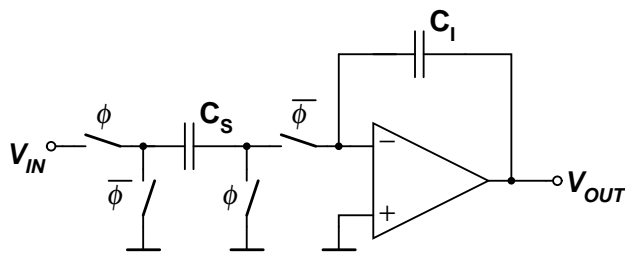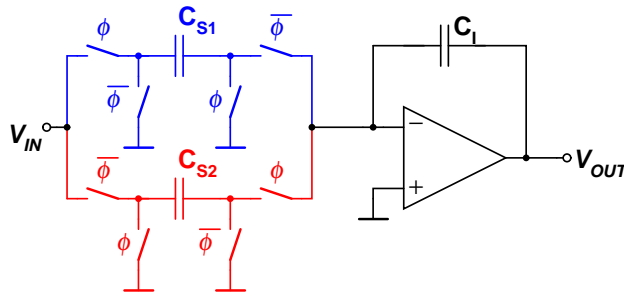
Figure 9.1 A switched capacitor integrator.

Figure 9.2 A double-sampled SC integrator.

9.1   Principle

An SC circuit can be divided into blocks, each comprising an opamp and a set of switches and capacitors. The SC integrator shown in Figure 9.1 can be used as an example block. The integrator operates in two phases; in the first phase the circuit samples its input, which is usually the output of some other block, in the capacitor C_S. The second phase can be called the integration phase or amplification phase. During it, the circuit performs a charge transfer from the sampling capacitor to the integration capacitor C_I with the aid of the virtual ground provided by the opamp.

Since the output of the circuit has to be fully settled by the end of the integration phase, it can already be sampled by the following circuit block in this clock phase. If this is done, the opamp is not needed in the sampling phase. Sometimes, the sampling phase is used to auto-zero the amplifier, i.e. to cancel its input offset voltage, but this is not necessary in many applications. Another possible way to exploit the opamp's idle phase is to duplicate the sampling circuitry and operate the two sampling circuits in opposite clock phases. In this way the opamp and the integration capacitance are shared between the two sampling circuits. The sampling rate of the resulted double-sampled circuit, shown in Figure 9.2, is twice that of the original circuit. There is, however, only a minor increase in power consumption, since it is dominated by the opamp, which typically uses the class A architecture and hence consumes power also when idle.

Double-sampling can also be applied to the S/H circuit shown in Figure 3.9, resulting in the circuit shown in Figure 9.3. During the first clock phase the input is sampled in the capacitor C_{S1} and the capacitor C_{S2} is connected to the feedback loop around the amplifier. In the next clock phase the roles of the capacitors are changed C_{S1} being in hold mode and C_{S2} in sample mode. The use of double-sampling in an S/H circuit has been demonstrated with the prototype presented in Section 12.2.
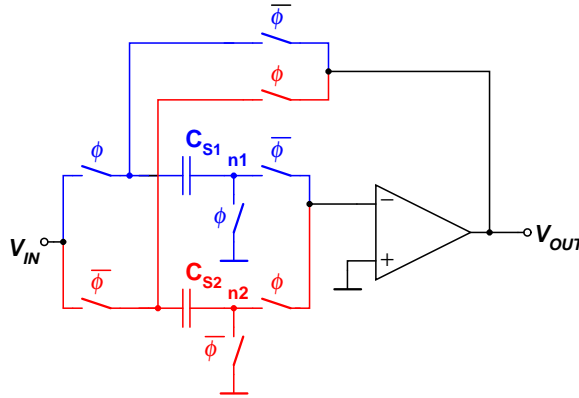
**Figure 9.3** A double-sampled S/H circuit.

9.2 Nonidealities

Double-sampling introduces some nonidealities not present in conventional SC circuits. Most of them arise from the mismatch between the two parallel circuits and are basically similar to the nonidealities in parallel ADCs, which have been analyzed e.g. in [121]. A general analysis concerning the timing of parallel sampling systems is given in [188]. Nonidealities in double-sampled SC filters have been investigated in [189]; however, not all the results are directly applicable to double-sampled S/H circuits. Thus, in the following sections the nonidealities of double-sampled circuits are analyzed, taking an approach better suited to S/H circuits and ADCs.

9.2.1 Memory effect

Due to the finite gain of the opamp a fraction of the previous sample remains stored in the parasitic capacitance in the input of the amplifier [96, 186]. Using the z-transform the voltage gain of the circuit in Figure 9.3 can be written as

$$\frac{V_{OUT}}{V_{IN}} = \frac{1}{1 + \frac{1}{A}\left(\frac{C_S + C_{in} + C_{p1}}{C_S}\right) - \frac{C_{in}}{AC_S}z^{-1}},\qquad(9.1)$$

where A is the DC gain of the opamp, C_S the sampling capacitor, C_{in} the opamp input capacitance, and C_{p1} the parasitic capacitance at node n1. The equation differs from the one for the conventional circuit (shown in Figure 3.9) in that it has an extra term, proportional to z^{-1}, in the denominator. The equation shows that double-sampling, together with the finite opamp gain and the parasitic capacitances, adds a low-pass
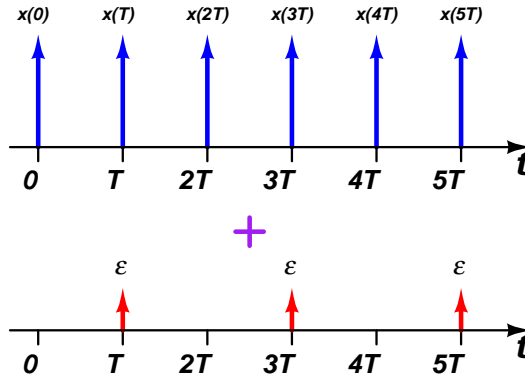
Figure 9.4 Time domain representation of a double-sampled signal in the presence of channel offset.

filtering effect. In the worst case (at the Nyquist frequency) the additional error is equal to the error caused by the opamp input capacitance in a conventional circuit.

9.2.2 Offset

A static DC offset between the two signal paths can be considered as a constant value added to every other sample. In the time domain this can be written as

$$y(t) = \sum_{n=-\infty}^{\infty} x(t) \cdot \delta(t - nT) + \sum_{n=-\infty}^{\infty} \varepsilon \cdot \delta(t - n2T - T), \tag{9.2}$$

where ε is the magnitude of the offset and $\delta(t)$ the Dirac's delta function. The time domain signal is illustrated in Figure 9.4.

The frequency domain representation can be obtained with the Fourier transform, which results in

$$Y(f) = \sum_{n=-\infty}^{\infty} X(f) \cdot \delta(f - \frac{n}{T}) - \sum_{n=-\infty}^{\infty} (-1)^n \varepsilon \cdot \delta(f - \frac{n}{2T}). \tag{9.3}$$

The equivalent magnitude spectrum is shown in Figure 9.5, where $f_S = 1/T$ is the clock frequency of the whole system, i.e. it is twice the clock frequency of the individual sampling circuits. The result obtained indicates that the offset between the two parallel signal paths results in tones at multiples of $f_S/2$.

In practice, large channel offset is not likely to occur in double-sampled circuits, since the opamp, which is the main source of offset, is common for both the signal paths.
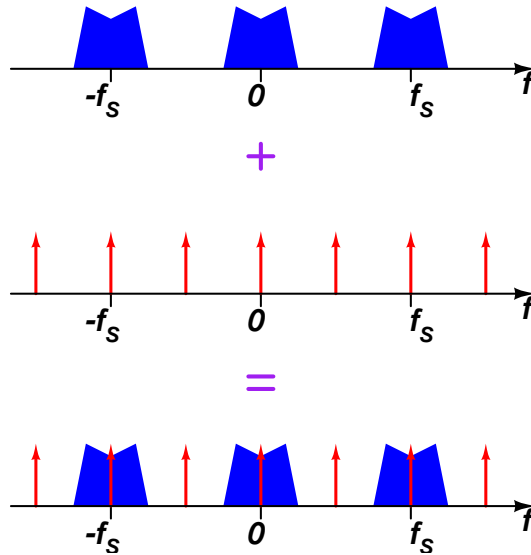
Figure 9.5 Frequency domain representation of a double-sampled signal in the presence of channel offset.

9.2.3   Gain Error

If there is a gain mismatch between the parallel circuits, the sample sequences they produce have different amplitudes. In the time domain this can be written as

$$
\begin{aligned}
y(t) \;=\; & 1 \cdot \sum_{n=-\infty}^{\infty} x(t) \cdot \delta(t - n2T) \\
& + (1-\alpha) \cdot \sum_{n=-\infty}^{\infty} x(t) \cdot \delta(t - n2T - T) \qquad (9.4) \\
\;=\; & \sum_{n=-\infty}^{\infty} x(t) \cdot \delta(t - nT) \\
& \cdot \left(\sum_{n=-\infty}^{\infty} \delta(t - nT) - \alpha \cdot \sum_{n=-\infty}^{\infty} \delta(t - n2T - T) \right), \qquad (9.5)
\end{aligned}
$$

where α is the normalized gain mismatch. Figure 9.6 shows that the gain mismatch in the time domain is equivalent to multiplying the ideal sample sequence by a sequence of two alternating constant impulses.

In the frequency domain the multiplication corresponds to the convolution and
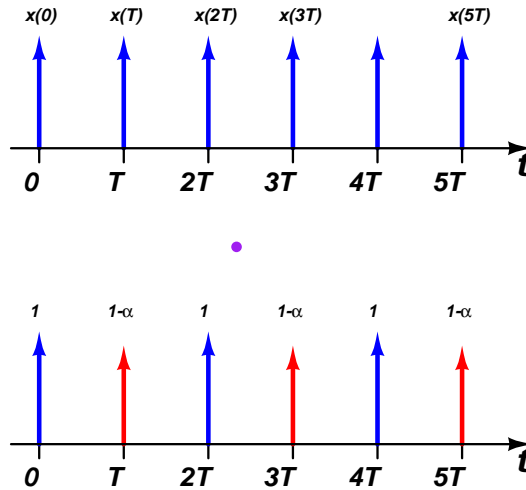
Figure 9.6 Time domain presentation of a double-sampled signal in the presence of gain mismatch.

thus the Fourier transform gives

$$Y(f) = \sum_{n=-\infty}^{\infty} X(f - \frac{n}{T}) - \alpha \cdot \sum_{n=-\infty}^{\infty} (-1)^n \cdot X(f - \frac{n}{2T}). \qquad (9.6)$$

This is illustrated in Figure 9.7, which reveals that the consequence of gain mismatch is parasitic sidebands around the multiples of $f_S/2$. If the signal bandwidth exceeds $f_S/4$ the sidebands alias to the signal band, degrading the signal-to-noise ratio. Even if the spectra do not overlap, filtering is needed to remove the sidebands.

Gain mismatch originating from capacitor mismatch is a severe problem in some double-sampled circuits. For example, in $\Delta\Sigma$-modulators the mismatch down converts the shaped noise energy around $f_S/2$ to the baseband [185]. However, in the S/H circuit shown in Figure 9.3 the gain is always one and independent of capacitor ratios, and thus gain mismatch is not a problem.

9.2.4   Timing Skew

There can be a constant timing skew in the clock signals of the two parallel circuits. This is illustrated in Figures 9.8 (a) and (b), where the sample sequences taken by each circuit are shown. The sequence $y_2'(t)$ has a constant timing error ΔT, and thus
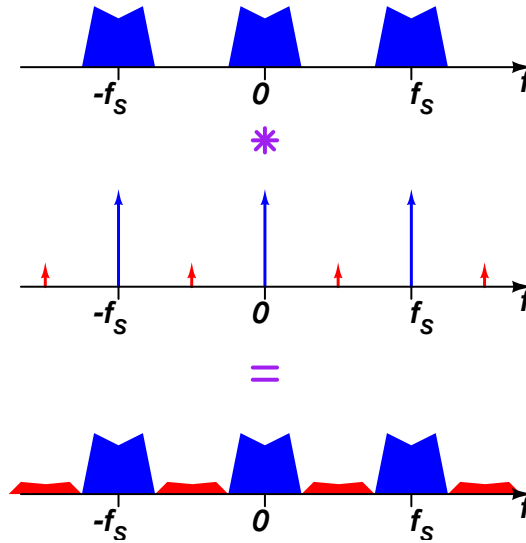
Figure 9.7 Frequency domain representation of a double-sampled signal in the presence of gain mismatch.

the sequences can be written

$$y_1'(t) \quad = \quad \sum_{n=-\infty}^{\infty} x(t) \cdot \delta(t - n2T) \tag{9.7}$$

$$= \quad \frac{1}{2T} \sum_{n=-\infty}^{\infty} x(t) \cdot e^{\frac{jn2\pi t}{2T}} \tag{9.8}$$

$$y_2'(t) \quad = \quad \sum_{n=-\infty}^{\infty} x(t) \cdot \delta(t - n2T - T - \Delta T) \tag{9.9}$$

$$= \quad \frac{1}{2T} \sum_{n=-\infty}^{\infty} x(t) \cdot e^{\frac{jn2\pi(t-T-\Delta T)}{2T}}. \tag{9.10}$$

The output of the circuit $y'(t)$ is the sum of y_1' and y_2'. In practical circuits it is held between the samples. If the subsequent signal processing is done in the discrete time domain, the held $y'(t)$ is resampled, resulting in a sequence of uniformly spaced samples, shown in Figure 9.8 (d). This sequence can be written as

$$y(t) \quad = \quad \sum_{n=-\infty}^{\infty} x(t) \cdot \delta(t - 2nT)$$

$$+ \sum_{n=-\infty}^{\infty} x(t + \Delta T) \cdot \delta(t - 2nT - T) \tag{9.11}$$
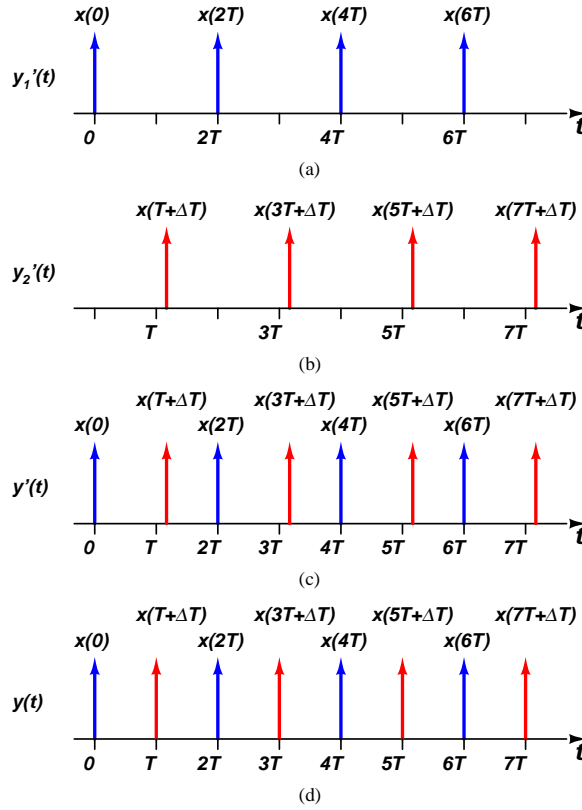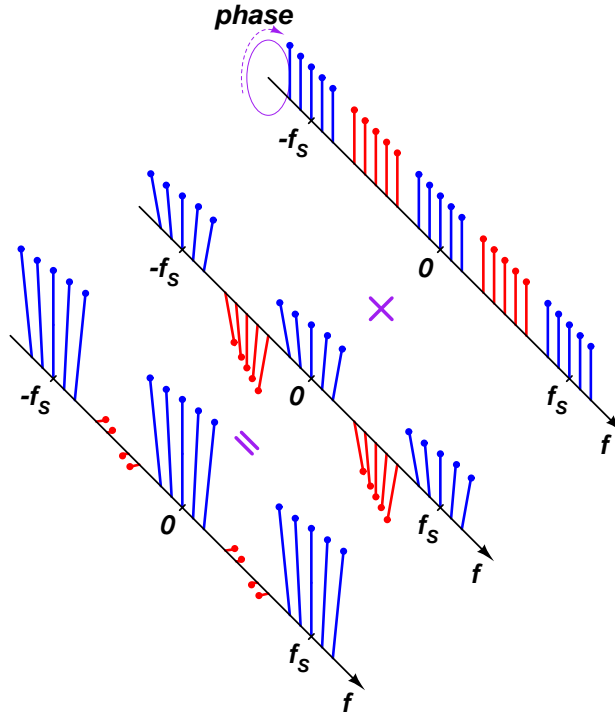
Figure 9.8 Time domain representation of a double-sampled signal in the presence of timing skew.

$$\begin{aligned}
= \quad & \frac{1}{2T} \sum_{n=-\infty}^{\infty} x(t) \cdot e^{\frac{jn2\pi t}{2T}} \\
& + \frac{1}{2T} \sum_{n=-\infty}^{\infty} x(t+\Delta T) \cdot e^{\frac{jn2\pi(t-T)}{2T}}.
\end{aligned} \tag{9.12}$$

The resampling thus corrects the sample misalignment but retains the incorrect sample values.

The frequency domain representation for the first sample sequence $y_1(t) = y_1'(t)$ (the sequence without the prime denotes the signal after the resampling) is

$$Y_1(f) = \frac{1}{2T} \sum_{n=-\infty}^{\infty} X(f - \frac{n}{2T}) \tag{9.13}$$

Figure 9.9 Frequency domain representation of a double-sampled signal in the presence of timing skew.

and for the second resampled signal

$$Y_2(f) = \mathcal{F}\left\{ \frac{1}{2T} \sum_{n=-\infty}^{\infty} x(t+\Delta T) \cdot e^{\frac{jn2\pi(t+\Delta T)}{2T}} \cdot e^{-\frac{jn2\pi(T+\Delta T)}{2T}} \right\}. \tag{9.14}$$

Using the time shift theorem, the Fourier transform in (9.14) yields

$$Y_2(f) = \frac{1}{2T} \sum_{n=-\infty}^{\infty} X(f - \frac{n}{2T}) \cdot e^{j2\pi f \Delta T} \cdot e^{-\frac{jn2\pi(T+\Delta T)}{2T}}. \tag{9.15}$$

The output signal is the sum of Y_1 and Y_2; thus,

$$Y(f) = \frac{1}{2T} \sum_{n=-\infty}^{\infty} X(f - \frac{n}{2T}) \cdot \left[1 + e^{-jn\pi} \cdot e^{j2\pi\Delta T(f-n/2T)} \right]. \tag{9.16}$$

These equations can be understood with the help of Figure 9.9. There, the magnitude and the phase of the signal spectra are represented in three dimensions—the

phase in the polar coordinates and the frequency axis in perpendicular to the phase plane. To make drawing easier, the base band signal spectra are represented as groups of impulses. The uppermost spectrum is $Y_1(f)$, which is the input signal sampled at the rate of $f_S/2$, and thus the baseband spectrum repeats at $f_S/2$ intervals. The spectrum in the middle is $Y_2(f)$ which, again, is the input signal sampled at $f_S/2$. The sampling moment, however, is ideally shifted by half a clock period compared to $y_1(y)$. The half period time domain shift causes the phase of the spectral images at odd multiples of $f_S/2$ to be rotated through 180 degrees. The samples, however, do not represent the input signal values at $nT + T/2$, but the time ΔT later. This rotates the edges of the spectral images in the phase plane.

Due to this bending, the sum of these two spectra (the bottom sequence) has some remnants of the spectral images at odd multiples of $f_S/2$, which would ideally ($\Delta T = 0$) be canceled out. If the bandwidth of the input signal is greater than $f_S/4$, these remains alias in the signal band. The wider the bandwidth, the larger the error signal, since the phase rotation is proportional to the frequency offset from the center of the image. Considering the time domain signal, this sounds reasonable, since the input signal change between the ideal and the actual sampling moment (i.e. the error) gets larger as the input frequency increases.

When the input is a sinusoidal signal (frequency f), the error is a tone at the frequency $f_S/2 - f$. When the magnitude of the error image is small compared to the fundamental, it can be approximated with $20\log(|\pi\Delta T f|)$ dBc, which is obtained from (9.16) using the small angle approximation.

The sources of the timing error are device mismatches in the clock generation circuit and uneven clock line capacitances. If the clock signals for the parallel circuits are generated using both the rising and falling edges of an external half-speed clock, the deviation of its duty cycle from 50% is seen as the timing error. These errors can be minimized but not totally eliminated by a careful layout design and by using a full-speed external clock.

9.3   Skew-Insensitive Circuit

To overcome the timing skew problem a modification, which makes the double-sampled circuits insensitive to the timing errors, has been proposed by the author in [9]. There, the idea is to perform the sampling with a single switch rather than two parallel switches. In Figure 9.10 the technique is applied in a double-sampled S/H circuit.

The sampling is performed by the switch S_0 clocked with the signal ϕ_S (Figure 9.11). The switches S_1 and S_2 act as a multiplexer, which controls the alternate
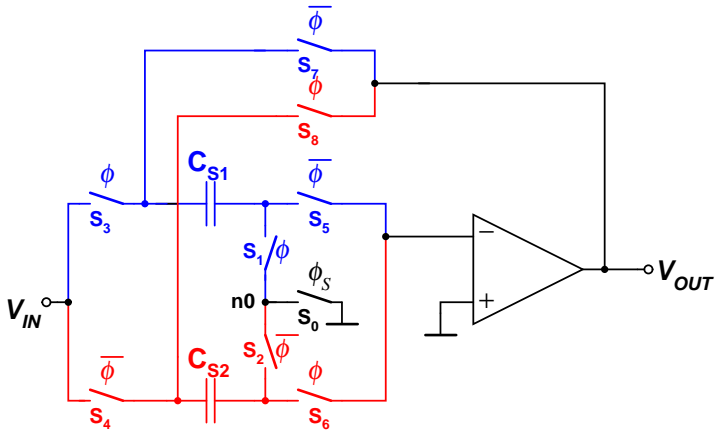
Figure 9.10 Proposed timing skew-insensitive double-sampled S/H circuit.
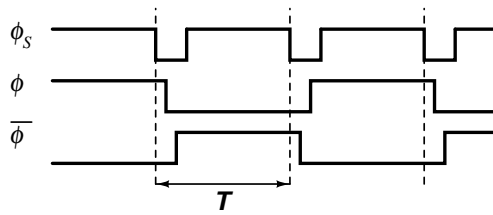

Figure 9.11 Timing of the skew-insensitive S/H circuit.

use of the common sampling switch in the parallel circuits. When the half circuit with the sampling capacitor C_{S_1} is in tracking mode switches S_0, S_1, S_2, and S_3 are on and switches S_5 and S_7 off. The sample is taken by applying a short zero pulse to switch S_0, during which switch S_1 is turned off, followed by S_3 being turned off. Next, switches S_5 and S_7 are closed connecting the capacitor C_{S_1} to the feedback loop around the amplifier.

The time gap between turning off S_0 and S_1 must be quite short, since the voltage at the floating node n0 changes along with the input voltage. The voltage change causes a part of the sampled signal charge to be distributed to the parasitic capacitance in that node. When switch S_1 is turned off the charge in n0 is isolated, distorting the total sampled charge. In addition to making the timing gap small, minimizing the parasitic capacitance in n0 also helps to diminish this error source.

Skew removal requires of the sampling pulse that it goes to zero before the multiplexer switch (S_1 or S_2) is turned off and that it remains zero until the turn-off has been completed. The pulse can be longer—even half of the clock period long—but the acquisition time of the sampling circuit becomes shorter as the pulse length is increased. However, if the circuit can track the input signal in half a clock period, a full-rate clock with a 50% duty cycle can be used, which simplifies the design of the clock generator.

This skew removal technique has been tested with the S/H circuit prototype presented in Section 12.3. Another similar technique has been developed (independently from the author) by Gustavsson and Tan [190].

Chapter 10

Switched Opamp Technique

As discussed in Section 2.6, the inadequate switch transistor gate overdrive is the main obstacle to the low-voltage operation of standard SC circuits. In Section 6.4.3 the bootstrapped switches were investigated as a solution for this overdrive problem. Their biggest disadvantage is the fact that even the simplest switch circuits may have more than ten transistors, which increases the area and complexity of a SC circuit, which contains dozens of such switches. The idea in the switched opamp (SO) technique is not to modify the switch but the circuit itself, so that the switches always have the maximum possible overdrive, which is $V_{DD} - V_T$.

10.1 Operation Principle

Let us look at the integrator implemented in the standard SC technique, shown in Figure 10.1. There, two types of switches can be identified: series switches (S_1 and S_5), which pass signals, the level of which varies over the whole voltage range, and shunt switches (all the rest), whose other terminal is connected to the analog ground potential. The switch S_4, although a series switch in a sense, falls into the latter category, since it operates against the virtual ground.

It is always possible to select a ground voltage equal to V_{SS} (or V_{DD}), which provides the maximum possible voltage overdrive for nMOS (or pMOS) switches. Similarly, the level of the virtual ground can be set freely, but it affects, of course, the design of the opamp input stage. The series switches are connected to the opamp output, where the signal common mode level is set to $V_{DD}/2$ to maximize the signal range. Consequently, the overdrive voltage, regardless of the switch type, is only $V_{DD}/2 - V_T$ in the worst case.
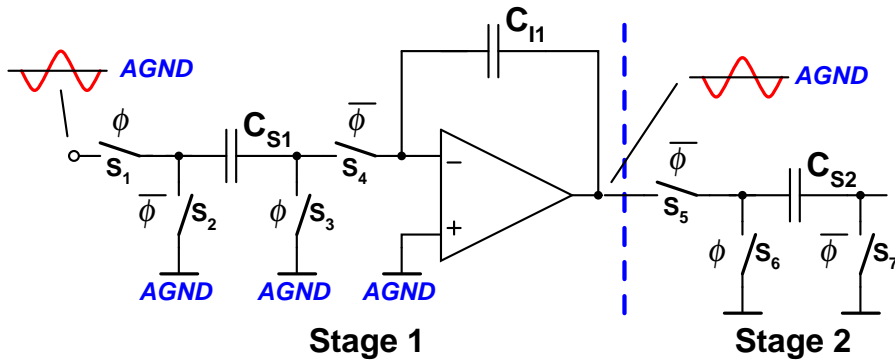
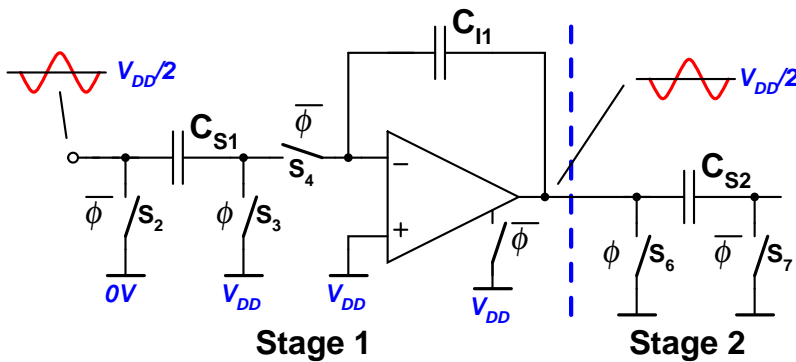**Figure 10.1** Switched capacitor integrator.

Figure 10.2 Switched opamp integrator.

The main idea in the switched opamp technique is to eliminate these series switches. The function of this type of switch, e.g. S_5 in Figure 10.1, is to disconnect the opamp from the next stage sampling capacitor C_{S2} in phase ϕ, when the capacitor's left terminal is shorted to the ground. If the opamp was not disconnected, there would be a race between it and switch S6. In an SO circuit this condition is avoided by turning the opamp output into a high impedance state during phase ϕ. Thus, like a tri-state logic circuit, the opamp does not resist the pulling of its output to the ground and so there is no need for a series switch. The SO implementation of the integrator, where the series switches are eliminated and the opamp made switchable, is shown in Figure 10.2.

The SO technique was first introduced in [191] and [36] and further developed in [192] by making the circuit fully differential and separating the input and output common mode levels. Reported SO circuits include filters [36, 193], ΔΣ modulators [194, 195], and the author's two pipelined ADCs [8, 2].

The rest of this chapter covers the implementation issues and the design of a
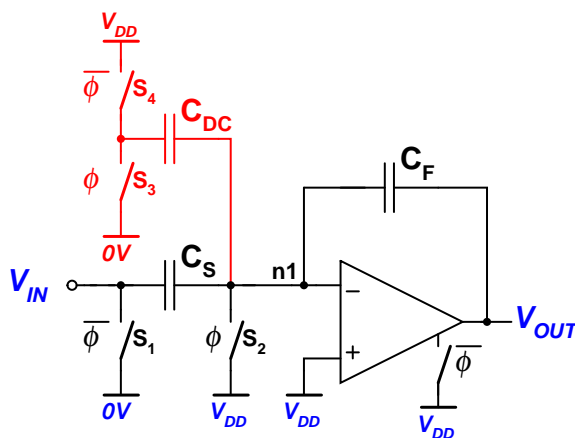
Figure 10.3 DC voltage correction with an extra capacitor C_{DC}.

switchable opamp. Interfacing the SO circuit to the outside world is discussed as well.

10.2   Compensating Common Mode Voltage Step

In SO circuits the signal DC level, or common mode level in fully differential circuits, is not the same in the two clock phases. The consequences of this and a method to avoid them are studied next.

Let us first ignore the capacitor C_{DC} in the SC amplifier of Figure 10.3 and assume that the signal AC voltage is zero. In the sampling phase ($\phi = 1$) V_{IN} is $V_{DD}/2$, V_{OUT} is 0 V, and node n1 is shorted to V_{DD}. In the amplification phase ($\phi = 0$) the situation in the input and the output is reversed; now $V_{IN} = 0$ and $V_{OUT} = V_{DD}/2$. When the capacitors C_S and C_F are equal, node n1 stays in balance, like the middle of a teeter board.

Usually, the capacitors are not equal, which results in an error, which is seen as an offset in single-ended circuits and a change in opamp input CM level in fully differential circuits. In [192] an extra switched capacitor (capacitor C_{DC} in the figure), which injects a constant correction charge into the node n1 every clock cycle, was proposed to overcome the problem. Setting the capacitor value to $(C_S - C_F)/2$ (assuming $C_S > C_F$) balances the voltage. In an integrator, the feedback capacitor is not reset and as a result the required C_{DC} is $C_S/2$.
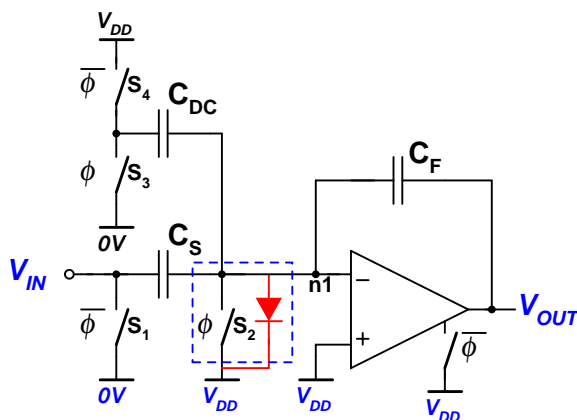
**Figure 10.4** Switch transistor junction diode may cause charge leakage.

10.3   Preventing Charge Leakage from Virtual Ground

The pn junctions at the drain and source of a MOS switch transistor form reverse-biased diodes to the transistor bulk, which is normally connected to V_{SS} in nMOS transistors and to V_{DD} in pMOS devices. If the signal voltage exceeds V_{DD} in a node where a pMOS switch is connected, the diode becomes forward biased and may leak some of the charge stored in that node. A similar situation occurs with an nMOS device when the voltage goes below V_{SS}. Typically, a momentary forward bias of less than 500 mV is not harmful with silicon diodes.

A node where the voltage may peak beyond the allowed limits and where the charge conservation is of the utmost importance, is the virtual ground in the opamp input. The situation in an SO amplifier which has the virtual ground (n1) set to V_{DD} is illustrated in Figure 10.4. The junction diode of the pMOS switch S_2 is shown between node n1 and V_{DD}.

In Figure 10.5 (a) the voltages in the input, the output, and the virtual ground are shown. The dotted curves represent the response for the minimum and maximum input signal voltages. The curves shown are for a circuit where C_S and C_F are equal. The switches and the opamp output stage are operated simultaneously. Consequently, due to the opamp's finite bandwidth, slew-rate, and the speed of the common mode feedback, the output voltage changes more slowly than the input voltage. As a result, the voltage at node n1 peaks safely downward. Charge injection from the switch S_2 causes upward peaking at the very beginning of the transient.

When the capacitor C_{DC} is present, it pushes the virtual ground in the opposite direction as C_S. The integrator is an extreme case, where the effect of C_{DC} on the DC
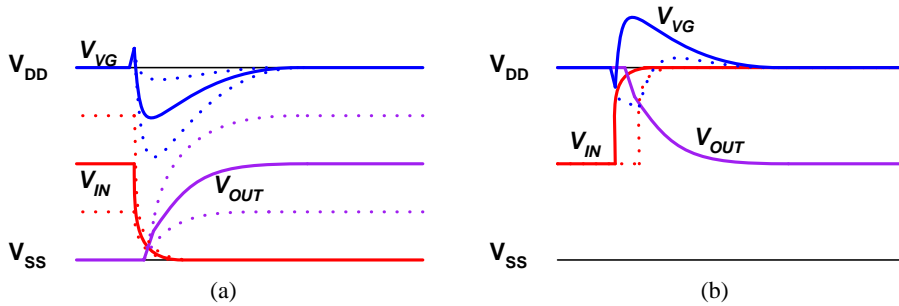
Figure 10.5 Voltages at the circuit input (V_{IN}), opamp output (V_{OUT}), and the virtual ground (V_{VG}) during a transient.

voltage is maximal. Assuming that the switched capacitors have equal time constants, the signal voltage may raise the virtual ground well above V_{DD} before the opamp responds and nullifies the voltage difference between its input terminals. The situation can be somewhat alleviated by ensuring that the voltage step caused by C_{DC} comes later than the one caused by C_S, either by making its time constant larger (by controlling switch on-resistance) or delaying the clock phase controlling the switching [193]. It has to be noted, however, that in very low-voltage circuits the signal voltages are also small, not easily causing spikes in excess the permissible $\sim$500 mV.

In an SO amplifier it is possible to set the output reset level and the virtual ground at the same potential (both to V_{DD} or V_{SS}). In an integrator forward biasing the diodes at opamp input during the sampling phase is not easily avoidable (it can, however, be done with yet another set of switched capacitors [196]), and thus the levels are normally different [195]. Figure 10.5 (b) shows again the voltages in the input, the output, and the virtual ground in an SO amplifier. Now charge leakage is a more serious risk, since the input step pushes the virtual ground in a hazardous direction. Delaying the switching of the input capacitors compared to C_{DC} and the opamp makes this configuration also usable as shown with dotted curves in the figure.

10.4 Speed

Achieving clock rates with SO circuits as high as with traditional SC circuits is not possible. There are several reasons for this, the most important of which is the finite opamp recovery time from the off-state. In addition, the limitation that the capacitors are permanently connected to the output of an opamp and the added extra capacitors lead, in many cases, to circuit topologies with a lower feedback factor than in SC
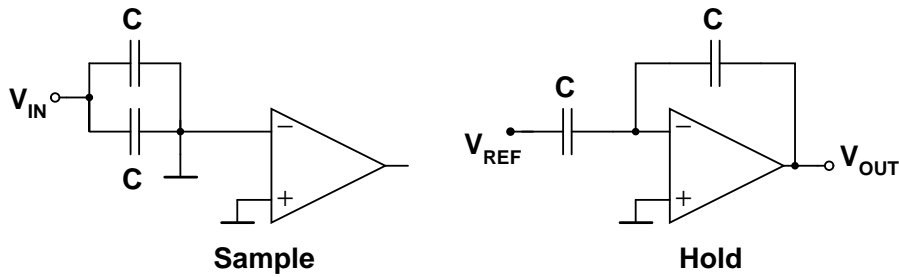
Figure 10.6 SC MDAC for a 1.5-b pipeline stage.

realizations.

Opamp recovery time depends on the settling of the internal nodes and charging the voltage step of $V_{DD}/2$ to the output capacitance. The internal settling can be improved by switching only the opamp output stage, disconnecting the compensation capacitors during the off-phase, and performing the switching with series switches between a current source transistor source and the power supply rather than a shunt switch at the current source transistor gate. In a differential SO circuit, switching of the opamp causes $V_{DD}/2$ common mode voltage step at the output. To minimize the recovery time the common mode feedback circuit has to be fast and the opamp slew rate toward the mid-supply level has to be high.

The penalty imposed by permanently-connected capacitors can be clearly seen when an SO MDAC is compared to its SC counterpart. Figure 10.6 shows the SC circuit typically employed in pipelined ADCs using the 1.5 bits/stage architecture. The input voltage is sampled in two equally-sized capacitors. In the hold phase one capacitor is used as a feedback capacitor while the other is connected to a reference voltage.

The SO implementation of the same circuit [2] is shown in Figure 10.7. Since the sampling capacitor cannot be disconnected from the previous stage, a separate feedback capacitor and capacitors for D/A conversion (subtracting the reference voltage) are needed.

The settling speed of the circuit is determined, besides by opamp bandwidth, also by the feedback factor of the circuit, which is given by the ratio of the feedback capacitor to the total capacitance. In the traditional circuit it is 1/2, while the larger number of capacitors in the SO circuit lowers it to 1/4. Expanding the analysis to a general MDAC, suitable for a k-bit pipeline stage, whose gain is $G (= 2^k)$, yields a feedback factor of $1/G$ for the SC circuit and $1/(3G/2+1)$ for the SO one.

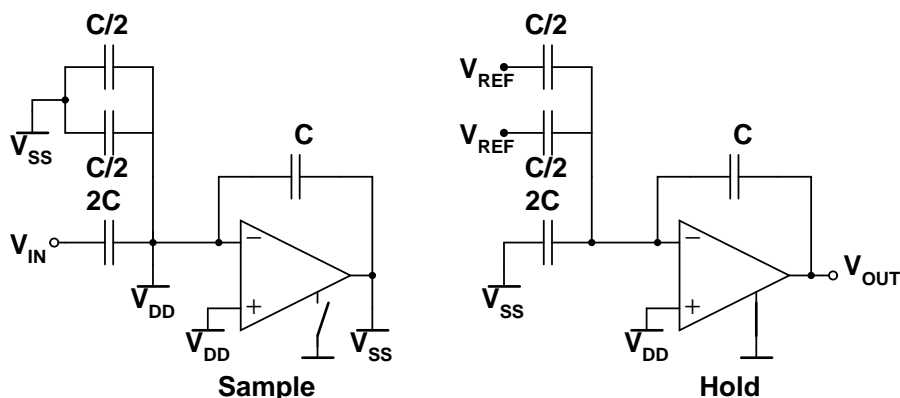One additional speed penalty of the SO technique is the lack of double-sampling

Figure 10.7 SO MDAC for a 1.5-b pipeline stage.

capability. Providing a valid output signal in both clock phases can, however, be accomplished by adding to the opamp a second parallel output stage, as the example, reported in [197], with a pseudo-2-path filter shows.

10.5   Power Supply Rejection and Noise

How switching the capacitors to V_{DD} and V_{SS} affects the power supply rejection ratio is an unavoidable question with SO circuits. It should be noted that in principle the voltages where the capacitors are connected need only to be at the same level as the supply voltages; no direct on-chip connection is required.

If the virtual ground level and the output reset level are equal (both V_{DD} or V_{SS}), this voltage level can be considered as a signal ground, against which all the signals are referred. For signal capacitors the situation is not different from that in traditional SC circuits. The extra capacitor C_{DC}, in contrast, is connected to the opposite supply rail, and the differential noise between the rails is coupled to the signal voltage, only attenuated by the capacitor ratio, which is typically of the order of 6 dB [193].

Alternatively, if the virtual ground level, which is also the level of the sampling ground, and the output reset level are opposites (one V_{SS} and the other V_{DD}, or vice versa), the situation is worse. Now the signal is sampled a against different voltage than the one against which the signal is referred in hold mode. As a result, the differential noise between these levels is directly summed to the signal voltage, i.e. the rejection is 0 dB.

Making the circuit fully differential ideally blocks all the noise that has been discussed. The rejection ratio is, of course, finite because of capacitor mismatch, but

should typically be at least 50 dB. Thus, even if the reference levels are the analog V_{DD} and V_{SS}, noise coupling through other routes is likely to dominate. In practice, all but the very first reported SO circuits have been fully differential implementations.

The extra switched capacitor C_{DC} adds its contribution to the thermal noise [195]. When its value is $C_S/2$, the increase in noise level is 1.8 dB, which can be compensated for by increasing the sampling capacitor by 50%.

10.6   Switchable Opamps

The switchable opamp is not essentially different from a conventional low-voltage opamp. In many cases an opamp can be made switchable simply by adding one or two transistors. Thus, the most important requirement for the opamp is a low-voltage capability. From the switch perspective the SO technique requires a supply voltage which is only V_T plus some overdrive. The same is expected from the opamp. A differential pair can operate with a $V_T + 2V_{dsat}$ supply, and hence it complies with the requirement. The other structures employed in the opamp need to be chosen in such a way that this limit is not significantly exceeded.

The choice of the common mode potential of the virtual ground and the type of the opamp input pair are linked; setting the CM level to V_{DD} requires nMOS input transistors, while setting it to V_{SS} calls for pMOS devices. If low $1/f$ noise is required, a pMOS input pair is preferred, while nMOS devices provide larger g_m leading to wider bandwidth and lower thermal noise. On the other hand, the choice of the input CM level also determines the type of the switches connected to the virtual ground, which has an effect on their on-resistance and parasitic capacitances.

The role of the common mode feedback circuit is emphasized in SO circuits for two reasons. First, the low voltage sets some limitations on applicable circuit structures and second, the change in output common mode level between the on and off phases requires high-speed common mode feedback capable of supplying high common mode slewing currents.

Finally, the opamp, or at least its output stage, has to be switchable. The most important requirement for the switching method is a fast recovery from the off state.

Next, some opamp circuits from the literature and three new proposals are studied in more detail.
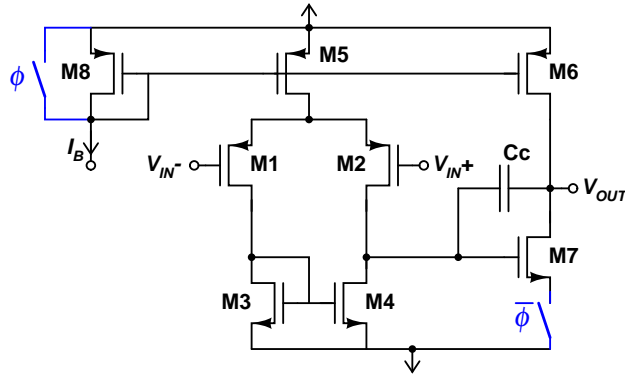
Figure 10.8 Steyaert's switchable opamp [36].

10.6.1   Circuits from the Literature

10.6.1.1   Steyaert's Switchable Opamp

The opamp in the original paper [36] by Steyaert and Crols is derived from the classic Miller topology and shown in Figure 10.8. The switching is implemented with two switches, one shunting the gates of pMOS current source transistors to V_{DD} and the other cutting off the current path between the output stage transistor M7 and V_{SS}. The circuit is single-ended and hence rather a proof of concept than a practical building block. Nevertheless, besides the biquad in the original paper, a $\Delta\Sigma$ modulator utilizing the opamp has been reported in [194]. The most prominent shortcoming of this circuit is the long recovery time resulting from the need to charge and discharge the gate capacitances of M5, M6, and M8 every clock cycle with a constant bias current I_B. The circuit has a minimum supply voltage of $V_T + 3V_{dsat}$ and it does not allow for the setting of the input CM level to V_{SS}.

10.6.1.2   Fully Differential Switchable Opamp

The fully differential opamp [198, 193] shown in Figure 10.9 improves the previous circuit in various ways. To begin with, the minimum supply voltage has been squeezed to $V_T + 2V_{dsat}$ by folding the first stage, which also allows for the setting of the input CM voltage to V_{SS}.

The recovery time has been improved by minimizing the effect of switching to the opamp internal nodes. Thus, the first stage is not switched at all and the output stage is only disconnected from the V_{SS} rail, since there is no need to disconnect it from V_{DD} to which the output is shorted anyway. The switching is done by cutting off the output
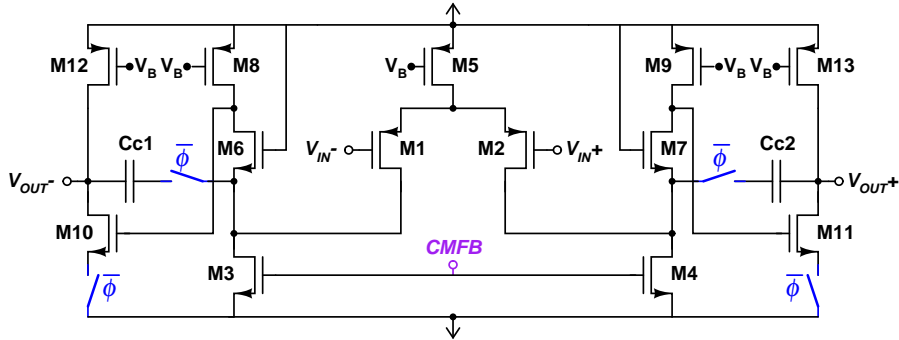
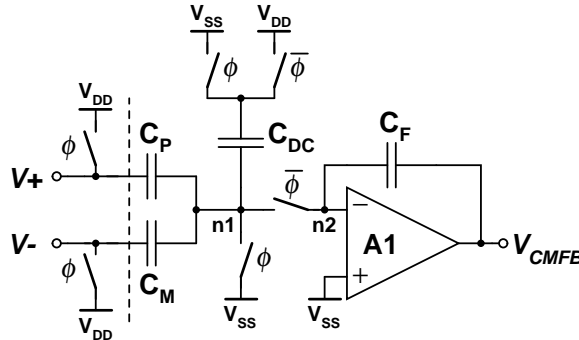**Figure 10.9** Fully differential switchable opamp [193].

Figure 10.10 Common mode feedback circuit for the opamp shown in Figure 10.9.

transistor current with a series switch, which leads to faster recovery than shorting the transistor gate to V_{DD}, because the first stage output is not disturbed and the voltage change over the gate-source capacitance is smaller. The compensation capacitors are disconnected during the off-phase to avoid discharging them.

A fully differential opamp needs a common mode feedback circuit. In a two-stage opamp the common mode voltages at the outputs of both the stages need to be controlled. This is most often accomplished by enclosing both stages in a single control loop, as also done in this opamp. The CMFB signal is applied to the first stage via the gate node of nMOS current source transistors M3 and M4. For the common mode signal the opamp has two cascaded inverting stages, and thus a negative feedback requires one extra inversion in the feedback path.

Since a series switch cannot be put in the output of the opamp, the traditional SC common mode feedback circuit [179] cannot be directly utilized in SO circuits. The circuit shown in Figure 10.10 performs the common mode sensing with a capacitive divider consisting of capacitors C_P and C_M, which are permanently connected to the
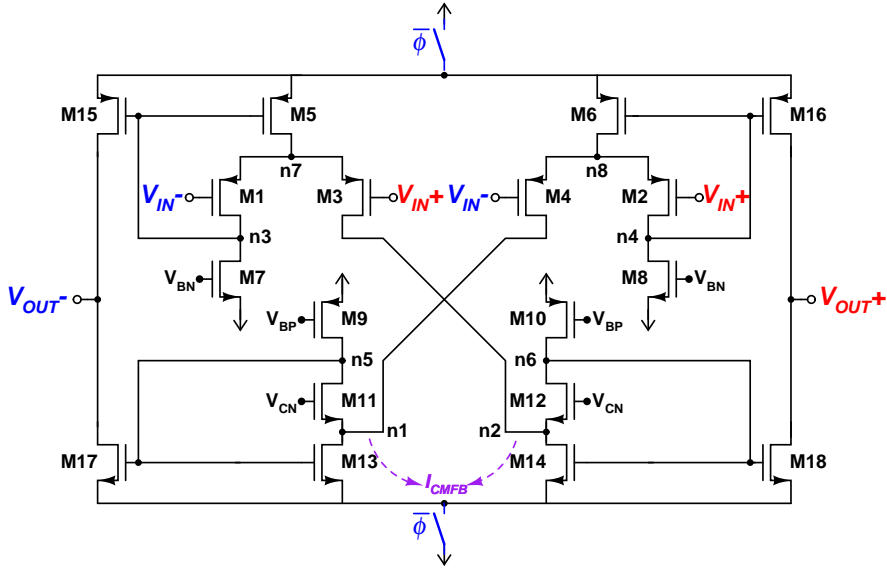
Figure 10.11 Class AB switchable opamp [199].

opamp outputs. The common mode voltage step is offset with the capacitor C_{DC}. As a result, node n1 is nominally equal to V_{SS} in both clock phases. The signal inversion and the generation of proper output DC level are performed by the feedback amplifier, consisting of a capacitor C_F and a single-ended opamp A1. (The circuit is actually an SC integrator, but here its transfer function is merely of interest in the continuous time domain, hence the term amplifier.) The circuit can be used without problems in low voltage applications, since all the switches are operated against V_{SS} or V_{DD} and the virtual ground is set to V_{SS}. It is, however, difficult to achieve simultaneously fast and stable CMFB because of the extra inverting stage in the feedback loop.

10.6.1.3 Class AB Switchable Opamp

A switchable opamp (Figure 10.11) based on class AB operation was proposed in [199] and later utilized in a $\Delta\Sigma$ modulator [195]. The fully differential circuit can operate with a $V_T + 2V_{dsat}$ supply and the class AB structure provides a moderate DC gain and high output driving capability. The common mode feedback is realized with a structure similar to the one in Figure 10.10, except that the SC amplifier is replaced with a simple open loop transconductor. The feedback signal is applied to nodes n1 and n2 in the form of two equal currents. Due to the single gain stage opamp architecture and the simplifications in the CMFB circuit, the resulting common mode
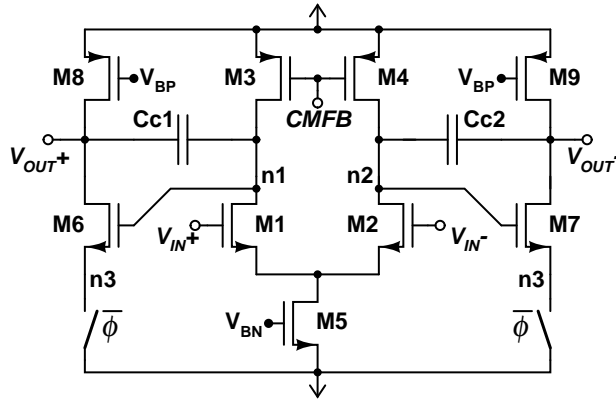
Figure 10.12 Fully differential switchable opamp based on the Miller topology.

settling can be made substantially faster than in the previous circuit. The switching is realized with series switches between the opamp local supply rails and the chips global supply rails.

The main disadvantage of the opamp is the fact that the circuit has practically only one gain stage, which, together with the large number of transistors, makes it noisy and sensitive to offset voltages. In addition, the low voltage current mirrors M5 and M6 restrict the minimum input common mode voltage $2V_T - V_{dsat}$ below V_{DD}. Consequently, there is a rather low maximum allowed supply voltage, which can be a limitation in applications requiring the capability of operating in a wide supply voltage rage.

10.6.2 Proposed Opamps

10.6.2.1 Opamp1

The first proposed switchable opamp has been developed for a pipelined ADC [8]. The circuit, shown in Figure 10.12, is based on a fully differential Miller opamp. To maximize the bandwidth and to minimize the needed supply voltage, the signal to the output stage is connected via the nMOS side. As a result the allowed supply voltage is in the range from $V_T + 2V_{dsat}$ to $2V_T + V_{dsat}$.

Only the output stage is switched. To prevent degeneration of the differential gain, the source nodes of M6 and M7 are actually connected together above the switches, although this is not shown in the schematic.

Figure 10.13 shows the common mode feedback circuit, which is based on the one shown in Figure 10.10. The signal inversion, however, is realized differently—here
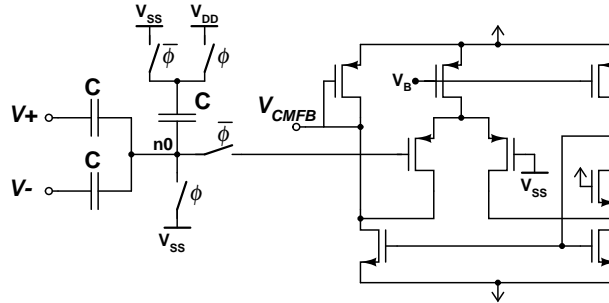
Figure 10.13 Common mode feedback circuit for the opamp shown in Figure 10.12.

with an open loop buffer instead of an opamp-based feedback circuit.

The main advantages of this opamp are its wide bandwidth, rather low thermal noise, and simplicity. On the other hand, the circuit suffers from a slow common mode slew rate, which is partly a consequence of the compensation capacitors not being disconnected during the off-phase. The capacitors cannot be disconnected, since their input stage side terminal is at a potential of $V_T + V_{dsat}$, which does not permit the realization of a well-conducting switch. As a result, when entering the on-phase, the common mode feedback, already not particularly fast, has to load the capacitors back to the nominal CM voltage, which slows it down considerably. Another drawback of this architecture is the limitation of the maximum supply voltage, which may be important in some applications.

10.6.2.2 Opamp2

The main problem with switchable opamps seems to be the recovery speed from the off state. In the next proposed opamp [6] the problem is tackled by increasing the bandwidth of the CMFB loop and by enhancing the common mode slew rate from the reset level toward the mid-supply level.

The opamp architecture, shown in Figure 10.14, is basically the same two-stage structure with a folded first stage as in Figure 10.9. The main difference is in the design of the common mode feedback circuit, where the main goal has been to get rid of the CMFB loop enclosing both the opamp stages, plus an extra inverter stage. This can be accomplished by having a separate CMFB loop for each stage. The first stage loop, however, is difficult to implement, and making the two loops settle nicely at the same time can be tricky. Thus, the implementation is based on a different idea, which is to totally eliminate the need for CMFB in the first stage and realize the second stage CMFB with a simple passive SC circuit, yielding a fast and stable single pole CM
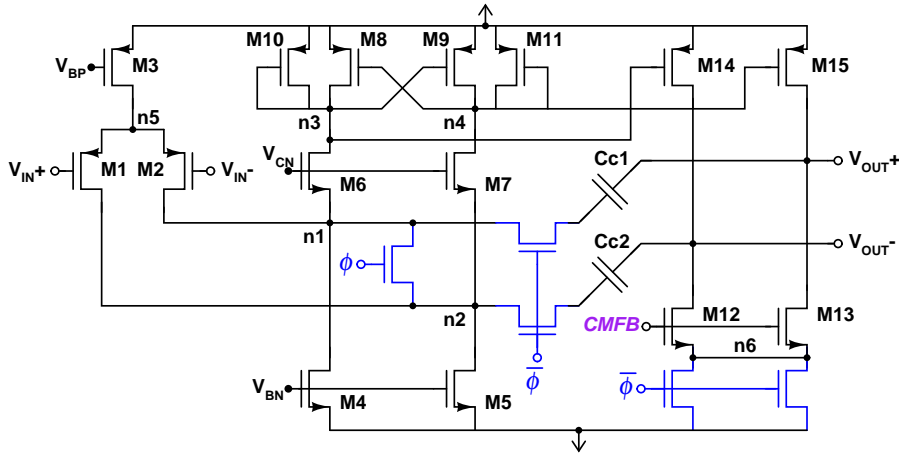
Figure 10.14 Switchable opamp with enhanced CMFB.

settling behavior.

The first stage is loaded, instead of simple current sources, with a structure commonly utilized in comparator preamplifiers [200]. It consists of the four pMOS transistors M8–M11, two of which (M10 and M11) are connected as diodes, while the other two are in parallel with them and have their gates cross-coupled to the complementary signal branch. Considering first a common mode signal, there is no difference between nodes n3 and n4, so they can be considered as shorted together. Consequently, the impedance seen from either of the nodes is approximately the parallel combination of the two transconductances, these being $1/(g_{m8} + g_{m10})$ for node n3. This is a low impedance, which yields a stable enough common mode voltage without any feedback!

When a differential signal is applied, the transconductance of the cross-coupled device appears with a negative sign and the equation, where the g_{ds} terms can no longer be neglected, becomes

$$r_3 = \frac{1}{g_{m10} - g_{m8} + g_{ds10} + g_{ds8}}. \tag{10.1}$$

Making the devices M8 and M10 identical and assuming perfect matching, the g_m terms in the equation cancel each other out, resulting in a high differential impedance, which is needed to achieve a high DC gain.

The potential problem in this type of load is the possibility that, for some reason, the impedance will become negative, making the opamp unstable. Mismatch between M8 and M10 is obviously one such reason, which is typically, however, small enough
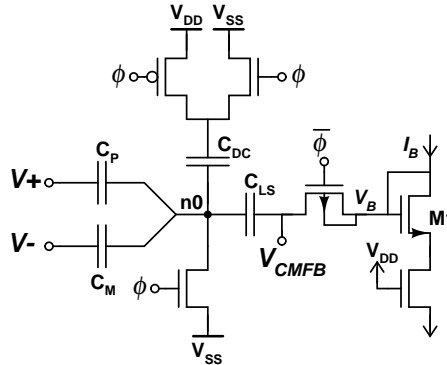
**Figure 10.15** CMFB for the opamp in Figure 10.14.

when the layout is drawn with care. Another, probably more serious threat is the dynamic mismatch, which is due to the fact that the g_m cancellation works only in small signal conditions. When the opamp is slewing, there can be a significant difference in the g_m's, which makes the impedance in one branch momentarily negative, causing oscillations in settling, which makes it considerably longer. Thus, when a high differential slew rate is required, this opamp is probably not the best choice. Methods for alleviating the potential problems include sizing the cross-coupled devices slightly smaller than the diode devices or using source degeneration. The slewing behavior can be improved by increasing the bias current.

For fast recovery, it is desirable that during the off-phase the fist stage output does not saturate in the presence of input offset, which is avoided by shorting nodes n1 and n2.

The common mode slew rate has to be high only in one direction, from V_{DD} to the mid-supply level in this circuit. The slew rate is maximized by having an active pull-down, realized with M12 and M13 and the CMFB, instead of a pull-down with a constant current.

Disconnecting the compensation capacitors in the off-phase serves two purposes. It enhances the common mode slew rate, since when connected back again the capacitors pull the nodes n1 and n2 up at the beginning of the settling. This works in the same direction as the common mode feedback and has a boosting effect on it. Furthermore, if the consecutive output signal values do not differ much, preserving the charge in the capacitors also reduces the differential settling time.

The CMFB circuit is shown in Figure 10.15. The sensing circuit, consisting of the capacitors C_P, C_M, C_{DC}, and the attached switches, is again similar to the one in Figure 10.10. Here signal inversion is not needed, only level shifting to a proper level,
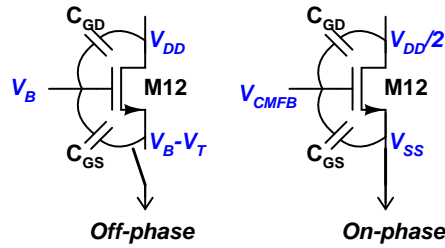
Figure 10.16 DC voltages at terminals of transistor M12 and its parasitic capacitances.

which is realized with the capacitor C_{LS}. The capacitor is precharged to the bias voltage V_B, which is generated with a bias current and a scaled-down replica of the opamp output current source. The voltage V_B is approximately equal to $V_{T,n} + V_{dsat}$, which is a high enough potential for a pMOS switch. The switch, however, has a rather high on-resistance as a result of the minimal overdrive, but it is not a problem, since, once reached, the purpose of the switch is only to maintain a constant precharge level in C_{LS}.

What makes the realization of the CMFB more complicated than it first looks is the fact that, due to the parasitic capacitances in the opamp output stage devices M12 and M13, setting the output CM level correctly is not straightforward. This is illustrated in Figure 10.16, where the voltages at the transistor terminals in both clock phases are shown. The voltage change disturbs the common mode feedback because of feedthrough via the parasitic gate-source and gate-drain capacitances, resulting in too high an output common mode level. The problem is reduced to a tolerable level by making the capacitors C_P and C_M large in comparison to the parasitics. Other possibilities include isolating the transistor gate in the off-phase with an additional series switch, using a dummy structure to compensate for the feedthrough, or buffering the CMFB signal with a continuous time buffer such as the one employed in opamp1.

The proposed circuit has since been used in a switched opamp $\Delta\Sigma$ modulator by Sauerbrey and Thewes [201], and a modified version of the circuit, where the cross-coupled load is in nMOS side and the output stage uses nMOS gain devices, in another low voltage $\Delta\Sigma$ modulator by Dessouky and Kaiser [145].

10.6.2.3 Opamp3

The third switchable opamp [2] is targeted to a pipelined ADC, where wide bandwidth, high slew rate, and relatively large DC gain are essential. In contrast, because of the
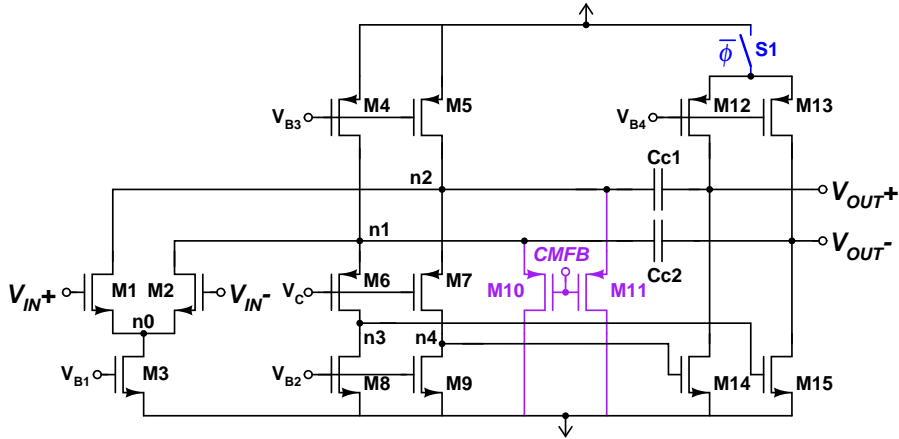
**Figure 10.17** Fully differential switchable opamp.

small feedback factor in SO MDACs, the phase margin requirement is not critical. Consequently, nMOS gain devices are utilized in both stages, yielding the architecture shown in Figure 10.17, which is practically an nMOS version of the opamp2 with a conventional current source load. The minimum supply voltage is again $V_T + 3V_{dsat}$.

The achievable DC gain is larger than in the previous two opamps utilizing a cascode first stage, since here the transconductance in both the stages is realized with nMOS devices. In addition, the extra V_{dsat} in the supply voltage, compared to the opamp shown in Figure 10.9, is over the transistors M8 and M9—just where it is most critically needed to improve first stage output impedance.

As discussed in Section 7.3.2.1, the combination of pMOS cascode devices and nMOS transconductances in the output stage is bad from the point of view of fast settling. When not aiming at unity gain stability, the situation is, however, manageable. Although the criterion (7.7) is not totally met, no peaking is observed, which is probably due to the stabilizing effect of the intrinsic Miller capacitance, formed by the gate-drain capacitances of devices M14 and M15.

The main reason for returning to a normal current source load is the requirement for a high differential slew rate, which is not easily obtained with opamp2 for the reasons given.

If the compensation capacitors were disconnected in the off-phase, the signal value from the previous clock period would remain on them. Since, in a pipeline ADC, voltage changes in opamp output can be from the smallest negative to the largest positive, this memory effect would increase the settling time in the worst cases. Thus, the capacitors are not disconnected, which also improves the phase response, since without
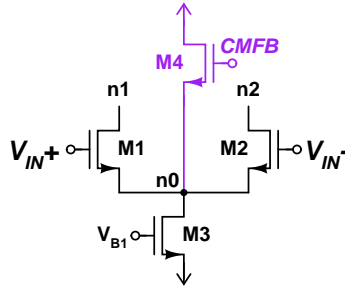
Figure 10.18 Alternative method to feed the CMFB signal to the first stage.

the switches there is no series resistance with the capacitors. The disadvantage of not disconnecting the capacitors, as in opamp1, is the increased demand for downward common mode slewing current in the opamp first stage.

The novel features of this circuit are again in the common mode feedback. The feedback signal is applied in the first stage, but for maximum speed and stability the number of extra nodes in the CMFB loop is minimized. Because of the need for signal inversion the sensed CM signal cannot be directly connected to the gate of any transistors in the first stage. Thus, the extra pMOS devices M10 and M11 are added to the cascode nodes and their currents are controlled with the CMFB signal. Alternatively, the same can be achieved with a single nMOS device connected to the node n0, as shown in Figure 10.18. The latter offers easier CM biasing (*CMFB* node biased nominally to V_{DD}) and less noise, but offers only limited slewing current for charging the compensation capacitors in contrast to the solution employed, which sinks a large downward current from the cascode nodes.

The common mode sensing is realized with the same structure as earlier (Figure 10.19); now, however, nothing else is needed. The capacitors are reset, instead of V_{SS}, against voltage V_C which is equal to the bias voltage for the cascode devices in the opamp. The voltage is low enough for the proper operation of an nMOS switch.

Returning the output common mode level to $V_{DD}/2$ from the reset voltage is a task to be performed every clock cycle. Since it is deterministic in nature, realizing it does not need to be totally based on feedback. Thus, the output stage current sources are dynamically biased with the structure shown in Figure 10.20. It uses a switched capacitor to push the gate bias downward at the beginning of the on-phase. This introduces a current pulse to the output stage, facilitating the task of the common mode feedback circuit. The size of the switched capacitor and the current I_B can be used to adjust the magnitude and the duration of the current pulse.
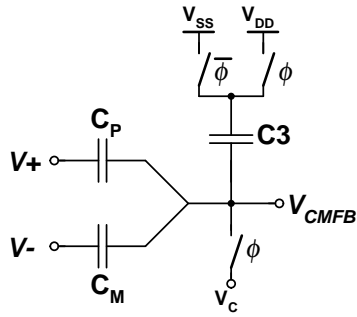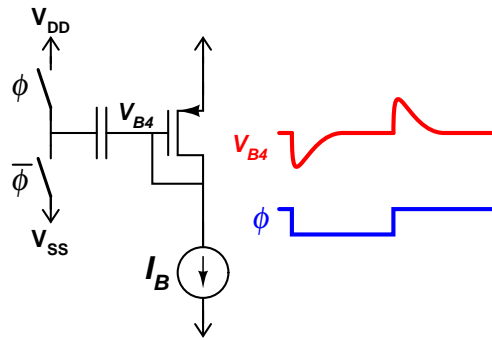
Figure 10.19 Passive CMFB circuit.

Figure 10.20 Dynamic CMFB boosting circuit.

Table 10.1 Switchable opamps compared.

	Baschirotto	Peluso	Opamp1	Opamp2	Opamp3
min supply	$V_T + 2V_{dsat}$	$V_T + 2V_{dsat}$	$V_T + 2V_{dsat}$	$V_T + 3V_{dsat}$	$V_T + 3V_{dsat}$
max supply	not limited*	$2V_T + V_{dsat}$	$2V_T + V_{dsat}$	not limited	not limited
GBW	$-$	$+$	$+$	$-$	$+$
phase margin	$+$	$+$	$+$	$+$	$-$
DC gain	$+$	$-$	$-$	$+$	$++$
thermal noise	$-$	$--$	$+$	$-$	$-$
slew rate	$+$	$++$	$+$	$-$	$+$
CMFB	$--$	$+$	$-$	$++$	$++$

* requires changing the cascode biasing from that presented, adding one V_{dsat} to the
minimum supply.

10.6.3 Switchable Opamps: Comparison

Table 10.1 shows a comparison of the presented switchable opamp topologies. It
should be noted that in most cases the GBW can be traded with the phase margin
by using the complementary topology (all nMOS devices changed to pMOS and vice
versa). What can be seen is the fact that all the topologies have their strengths and
weaknesses, making the choice of the opamp architecture application specific.

10.7 Input Interfaces for SO Circuits

Switching the opamp is a way to get rid of the series switches connected to the opamp
output. Typically, however, there are also series switches in the circuit input, which
are not eliminated by the technique.

In Steyaert's and Crols' SO circuit [36], which was a low pass biquad, two ap-
proaches were tested. The first one was a switch controlled with a voltage higher than
the supply, which is, however, not a true low-voltage technique. The same, using a
long-term-reliable bootstrapped switch, has later been proposed in [136]. The second
solution was to implement the first resistor, normally realized with a switched capac-
itor, as a real resistor. This technique can be used in some SO filters, but it is not
applicable to all SO circuits.

Reduced swing input signal in conjunction with a series switch was used in the
$\Delta\Sigma$ modulators presented in [194] and [195]. This is not a true low-voltage technique
either, since when the supply voltage is reduced to a level where all switches are still
operational the available input signal swing is zero.

Bandpass circuits often have a non-switched capacitor in their input, and thus they
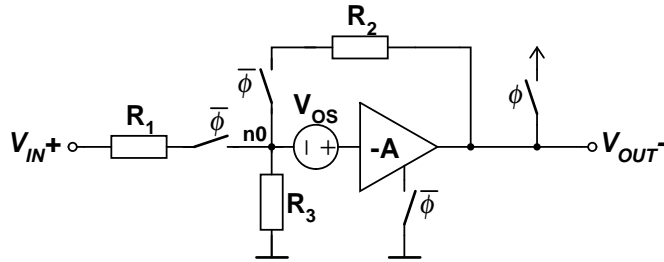can be implemented in the SO technique without problems [193, 202].

Figure 10.21 Input structure based on feedback amplifier.

10.7.1   Active Input Structures

In Nyquist rate A/D converters DC decoupling is out of the question and reduced signal range is unacceptable because of thermal noise and increased accuracy requirements for the comparators. At the time when the first SO pipeline ADC (Section 12.7) [8] was designed, no suitable solutions for the input circuitry existed.

The structure employed, shown in Figure 10.21, is based on a continuous time feedback amplifier. It makes possible the moving of the series switch from the circuit input to the virtual ground created by the opamp, making low-voltage operation possible. Recently, similar techniques for bringing the opamp input near supply rails have been investigated in conjunction with continuous time low-voltage circuits [203, 204].

The opamp is switchable, looking like just another switchable opamp to the following circuitry. In the on-state the DC voltage level at the amplifier output is $V_{DD}/2$ and the node n0 is near the ground level. Biasing the virtual ground to a level different from the output level is accomplished with a level-shifter (voltage source V_{OS}) and the extra resistor R_3. Instead of the resistor, a current source can be used, which avoids the degrading of the opamp gain but does not permit the biasing of n0 as low as the resistor and makes the biasing less robust. Furthermore, the current source adds more parasitic capacitance than the resistor.

The lower the voltage at n0, the better the switches conduct. In contrast, low n0 requires small R_3, which degrades the effective gain of the opamp, as seen from the transfer function, which is given by

$$\frac{V_{OUT}}{V_{IN}} = -\frac{R_2}{R_1} \cdot \frac{1}{1 + \frac{1}{A}\left(1 + \frac{R_2}{R_1} + \frac{R_2}{R_3}\right)}. \tag{10.2}$$

The circuit is intended to allow signal frequencies in the megahertz range. Thus, the bandwidth of the opamp is more important than the DC gain. Consequently, the
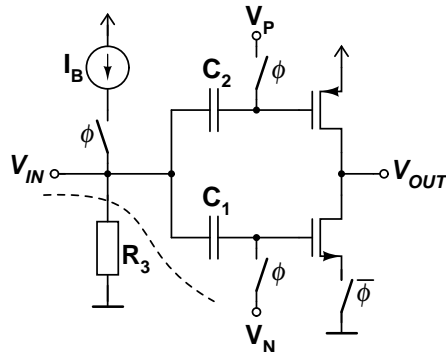
Figure 10.22 Implementation of the amplifier and the voltage source of Figure 10.21.

opamp is a simple inverter, shown in Figure 10.22. It operates in class AB, which is achieved via dynamic biasing, realized with switched capacitors, which also form the offset voltage source (V_{OS} in Figure 10.21). The capacitors C_1 and C_2 are refreshed during the off-phase by connecting their right terminals to the gate bias voltages of the transistors. The bias voltage on resistor R_3, which is also the second bias voltage for the capacitors, is produced with a switchable current source I_B.

The main drawback of this input structure is its limited linearity, which is due to inadequate opamp gain at the signal frequency and the signal dependent on-resistances of the series switches, which are not negligible compared to the resistor values.

While the ADC prototype was still in the process of fabrication, a similar type of input structure was proposed by Baschirotto et al. in [205]. The circuit is shown in Figure 10.23. It is targeted on lower signal frequencies, allowing larger resistor values, which makes it possible to leave out the series switches and simply ground node n0 in the off-phase without a fear of input signal feedthrough. As a result, n0 can be biased to $V_{DD}/2$ and no extra resistor or current source is needed. The opamp utilized is a two-stage structure providing good low frequency linearity. This circuit also suffers from limited amplifier bandwidth, which degrades the high frequency linearity.

Input structures utilizing a transimpedance amplifier instead of an opamp have recently been studied in [206]. These circuits have potential for somewhat larger bandwidths than opamp-based circuits.

The author also initially started looking for partially current mode input structures [207], but rejected the approach in favor of a totally passive circuit.
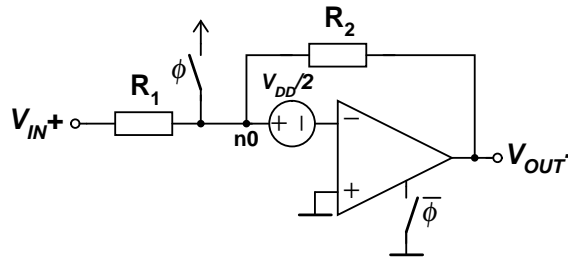
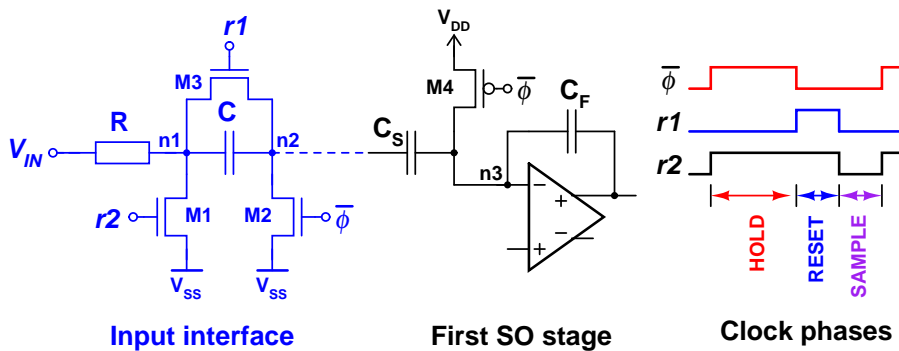**Figure 10.23** Another active input structure [205].

Figure 10.24 Passive input interface.

10.7.2 Passive Input Interface

The input structures presented are based on the utilization of an active circuit block, typically an opamp, whose finite bandwidth limits their linearity at high signal frequencies. Hence, in high frequency applications a circuit without an opamp would be attractive. Such a circuit was developed for the second ADC prototype (Section 12.8) [2]. It is based on the idea that a DC decoupled signal can be brought into SO circuits without problems, and since the opamp input is purely capacitive, DC decoupling only leads to the loss of signal DC value. Even this can be avoided, if the DC voltage on the coupling capacitor is known. Controlling the voltage on the capacitor can be accomplished by resetting it every clock cycle.

A circuit realizing this idea, together with the required clock signals, is shown in Figure 10.24. The input interface comprises resistor R, coupling capacitor C, and switch transistors M1-M3. The first SO stage is partially shown on the right of the input structure. Figure 10.25 shows simulated waveforms obtained with a 1.7-MHz signal at a 5-MHz clock rate.

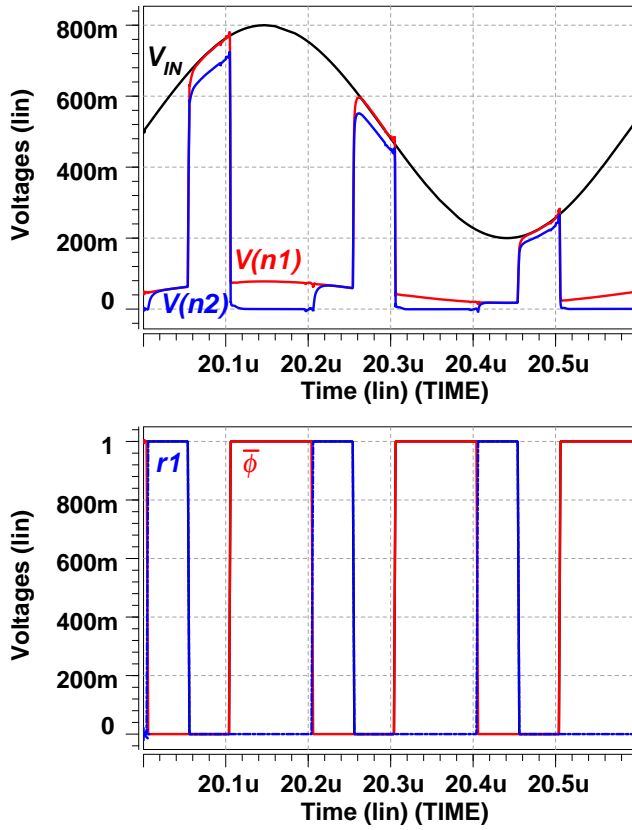The circuit uses three clock phases, the hold phase lasting a half clock period,

Figure 10.25 Simulated voltages in the circuit shown in Figure 10.24.

while the reset and the sample phase are each a quarter period long. During the reset phase ($\bar{\phi} = 0$, $r1=1$, $r2=1$) the voltage on capacitor C is reset with shunt switch M3. To allow the operation of the n-type reset switch node n1 is shorted to the ground and node n2 is left floating.

In the sampling phase ($\bar{\phi} = 0$, $r1=0$, $r2=0$) the input voltage is sampled in the series combination of the coupling capacitor C and the sampling capacitor C_S. Making C large compared to C_S results in the majority of the signal appearing across C_S. The remaining attenuation can be compensated for by properly adjusting the ratio of C_S and C_F in the first SO stage by setting

$$C_S = C_S' \cdot \frac{C + C_2}{C - C_S'},$$

(10.3)

where C_S' is the nominal value of C_S and C_2 the parasitic capacitance at node n2. The remaining gain error resulting from the uncertainty of C_2 is small enough for most applications.

In the hold phase ($\bar{\phi} = 1$), which is also the on-phase of the switchable opamp, both nodes n1 and n2 are shorted to the ground and the reset switch is open so as to improve isolation against input signal feedthrough. To the first switched opamp stage the input circuitry looks just like another switchable opamp, when being in the off-state. In this phase the charge sampled in C_S is transferred to the feedback capacitor C_F.

The size of resistor R is set by two constraints. First, it should be large in order to minimize the fractional signal voltage seen at n1, which results from resistive division between R and M1's on-resistance and causes signal feedthrough in the hold phase. For the same reason, M1 and M2 should be wide devices. Their size, however, cannot be made arbitrarily large, since their nonlinear parasitic capacitances are a source of harmonic distortion. On the high side the size of R is limited by the sampling time constant. The size of C is limited by the available area and its bottom plate parasitic capacitance, as well as the time constant associated with the resetting.

There is one potential problem in the circuit, which can, however, be avoided by one additional switch; when the capacitor is being reset nodes n1 and n2 are shorted and as a result an attenuated version of the input signal is seen at node n2: $V_{2R} = V_{IN} \cdot R_{ON1}/(R_{ON1} + R)$. At the end of the reset phase this voltage is sampled in C_S, introducing an error to the signal voltage, which will be sampled in the next phase. The error is equal to

$$V_{2S} = V_{2R} \cdot \frac{C_S + C_2}{C + C_S + C_2}.$$

(10.4)

The effect on the frequency response can be modeled with a two-tap FIR structure. If the effect is intolerable, which is unlikely, it can easily be almost totally eliminated by adding a series switch on the opamp side of C_S, which enables it to be disconnected during the reset phase.

The input structure does not significantly increase thermal noise, since the noise sampled in the sampling phase is the normal kT/C_S and the additional noise sampled in the reset phase is determined by the total capacitance, including the large capacitor C.

The clock signals with 25% and 75% duty cycles are most easily realized by using a double rate clock signal, from which all the necessary clocks are generated.

Chapter 11

Other Low Voltage Techniques and Building Blocks

11.1 Low Voltage SC Technique with Unity-Gain-Reset Opamps

The main factor which makes switched opamp circuits slower than traditional SC circuits is the time it takes to wake up the opamp from the off-state. Opamp switching is needed to make possible the shorting of the output node to ground, which is in turn needed by the charge transfer occurring in the following stage. The charge transfer, however, does not require the output reset level to be either of the supply levels; it can be any constant voltage.

The idea proposed in [208] uses this by connecting the opamp into unity gain feedback instead of turning it off and shorting the output to ground. Consequently, the opamp output settles to the virtual ground level in the reset phase. The principle of this technique is illustrated in Figure 11.1, where two cascaded integrators are shown.

Originally it was proposed to set the reference level V_{REF} equal to V_{SS}, which permits a supply voltage as low as in the switched opamp technique. This, however, introduces two problems. First, the switch on the left side of the integration capacitor C_3 is an nMOS transistor, whose junction diode easily becomes forward biased when the signal voltage on the capacitor pushes the node between the switch and the capacitor down in the reset phase. Second, driving the opamp output all the way down to V_{SS} in the reset phase pushes its output stage transistors out of saturation, resulting in
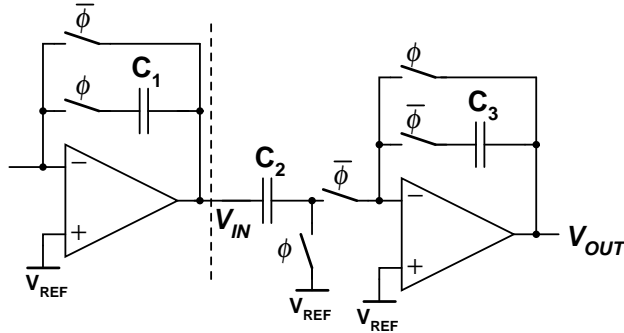
**Figure 11.1** SC circuit based on unity-gain-reset opamps.
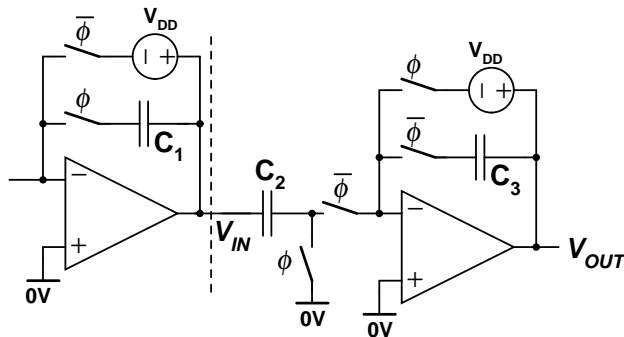
Figure 11.2 Floating voltage source in the feedback loop prevents the charge leakage from integration capacitor.

a recovery time which is not much better than in a well-designed switchable opamp.

In [209] a fully differential circuit was designed using V_{REF} of 500 mV, which does not permit as low a supply voltage, but removes both the problems mentioned. Furthermore, letting the output CM level be the same 500 mV also in the integration phase removes the need for extra DC correcting capacitors.

The original paper [208] proposes adding a floating voltage source in the unity gain feedback loop so as to prevent the charge leakage without increasing the supply voltage. The resulting circuit is shown in Figure 11.2. Now the output is reset to V_{DD} instead of V_{SS}, which pulls the node behind the integration capacitors up when entering the reset phase, and thus no leakage can occur. The opamp output is still driven out of saturation, which can be avoided by making the voltage source somewhat smaller than V_{DD}. The voltage source can easily be realized with a switched capacitor, as shown in the paper. The same authors have demonstrated the feasibility of the technique via the design of a $\Delta\Sigma$ modulator reported in [210].
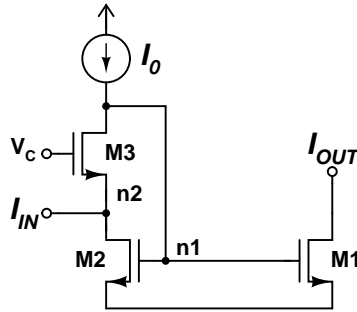
Figure 11.3 Low voltage current mirror.

Different CM levels in the reset and integration phase still require an extra switched capacitor, just like SO circuits, for adjusting the CM level. Therefore, I propose here that the voltage source should be $V_{DD}/2$, which removes the need for the capacitor while offering maximal voltage swing with minimal supply. The implementation of a voltage source is, however, not easily done with a switched capacitor, but can be realized with a resistive level shifter [211].

The non-switched opamp makes the technique potentially faster than the SO one. However, the capacitors are still permanently connected to the opamp outputs and the opamp is not available for signal processing during one half of the clock period. Thus, the technique still features some speed penalties in comparison to SC circuits. Further, the settling of a single stage involves two opamps, making it potentially longer. The resetting also requires the opamp to be unity gain stable with a good phase margin, which is not generally required in SO and SC circuits.

11.2   Current Sources and Mirrors

The traditional current mirror, either a simple topology or a cascoded structure, sinks the current through a diode-connected transistor. As a result the input voltage is $V_T + V_{dsat}$, which does not leave much room for other circuit structures in low voltage realizations. The low voltage current mirror, shown in Figure 11.3, removes this limitation by using the cascode node as the input for the signal current; only the DC bias current (I_0) is supplied in the traditional way. Thus, the minimum input voltage is now only V_{dsat}.

Since the signal current does not flow through the cascode transistor M3, the voltage variation at node n2 is very small (V_{n1} divided by the gain of M3), i.e. the circuit has a very low input impedance. This property can be exploited to realize a linear
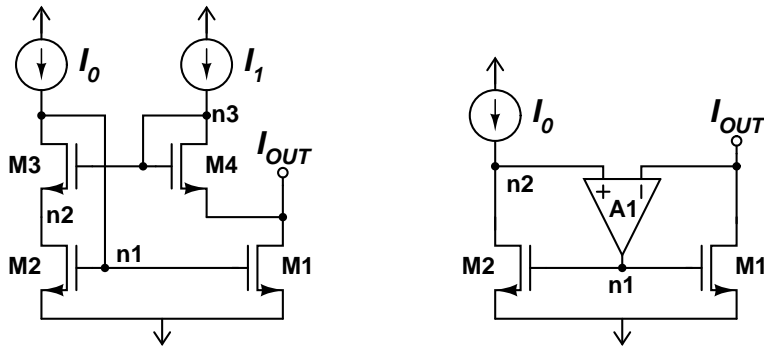
Figure 11.4 Triode region current source—two implementations.

voltage-to-current converter by placing a series resistor in the input of the current mirror [212].

Even a lower input, as well as output, voltage can be realized by biasing the transistors M1 and M2 in the triode region instead of saturation. Then, however, the output impedance becomes so low that the circuit no longer acts as a current source (or mirror). This can be corrected by adding a feedback loop, which makes the voltage at node n2 track the output voltage. Two realizations are shown in Figure 11.4. The one on the left [213] controls the gate voltage of the cascode transistor M3 with a level shifter constructed with a diode-connected transistor M4, which is matched with M3. The circuit on the right [214] uses an opamp to form the feedback loop. Another similar type of circuit is reported in [215].

In the literature [214, 215] the triode region current source has been employed as a tail current source in opamps. In principle it can be utilized in the output stage as well, but the local feedback loop has an effect on the opamp frequency response, requiring at least careful settling analysis and simulations.

11.3   Bandgap References

Like many other analog circuits, ADCs need a reference voltage, which is used for determining the quantization levels. In experimental prototype circuits the reference voltage is often supplied externally, while many commercial devices have an on-chip voltage reference, which is more convenient for the end user. The on-chip reference is typically realized with a bandgap reference (BGR) circuit, which provides a stable reference voltage over a wide temperature range.
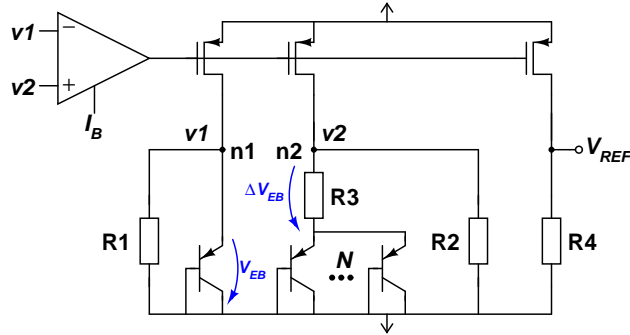
Figure 11.5 Low voltage bandgap reference.

11.3.1   Low Voltage BGR Circuits

The operation of bandgap reference is based on the fact that the base-emitter voltage (V_{BE}) of a bipolar transistor has a negative temperature coefficient, while in the voltage difference of two base-emitter junctions (ΔV_{BE}), biased with different current densities, the temperature dependency is positive. Thus, a properly weighted sum of them ($V_{BE} + K \cdot \Delta V_{BE}$) is free of temperature dependency. In traditional BGR circuits the sum is formed in the voltage domain, resulting in a reference voltage around 1.25 V, which is clearly an obstacle to low voltage operation.

A lower supply can be used if currents are summed instead of voltages. Then the required minimum supply voltage is the voltage of a forward-biased base-emitter junction plus a headroom for a current source. Thus, the circuit can be realized with a supply voltage around 0.9 V. Such circuits have been reported in [216, 217, 218, 219].

The low voltage BGR proposed in [217] is shown in Figure 11.5. The feedback loop, consisting of an opamp and a pair of matched controlled current sources, forces the voltages *v1* and *v2* to be equal. Consequently, the current through the resistor R1 is proportional to V_{EB} and the current through the resistor R3 to the difference of the emitter-base voltages of the two pnp transistors. Setting the resistor R2 equal to R1 makes their currents the same. Since the current of the controlled source is the sum of currents through R2 and R3, it will be proportional to $V_{EB} + K \cdot \Delta V_{EB}$, which is exactly what is required from a temperature-independent reference. The current generated is mirrored through the resistor R4, producing the reference voltage across it.

The biggest problem in this circuit is the realization of the opamp input stage. With the temperature, the emitter-base voltage (*v1* in the circuit) goes to about 500 mV at the lowest and above 800 mV at the highest. Thus, without increasing the supply voltage, there is not enough room for a MOS transistor gate-source voltage between
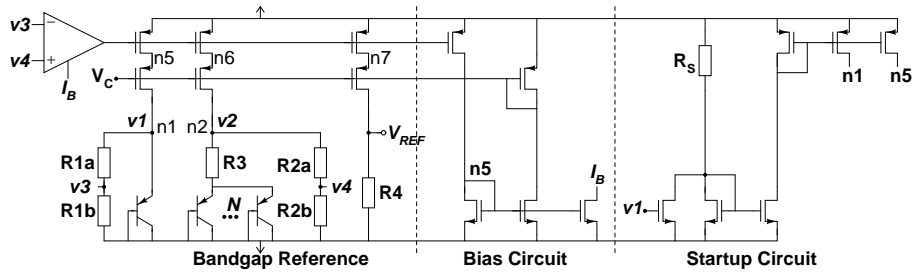
Figure 11.6 Proposed low-voltage BGR circuit with startup.

v1 and either of the supply voltages. In the future, when the MOS threshold voltage is scaled down close to 300 mV, the problem will disappear, which will also be the case when the process offers low V_T MOS transistors. In a BiCMOS technology, similar npn transistors which are used for generating the bandgap voltage can be used as the opamp input devices [219].

Two solutions for a standard CMOS technology have been proposed. One is to replace the opamp with a transimpedance amplifier and connect the resistors R1 and R2 to its inputs instead of the ground [218]. And the second, proposed by the author [14], is shown in Figure 11.6. There, the opamp uses a pMOS input pair and its inputs are connected to intermediate taps of resistors R1 and R2. The tap voltages *v3* and *v4*, which are in a 1:3 proportion to the voltages *v1* and *v2*, are low enough for the pMOS input pair.

The current sources are cascoded to increase output current accuracy. This, unfortunately, eats the precious voltage headroom needed for low noise and low offset biasing. With low V_T transistors it would be possible to bias the output cascode in such a way that the voltage at node n7 tracks *v1*, which would make the cascodes in the other two current sources redundant.

The BGR circuit has two stable operation points: the desired one and the other where the current is zero, i.e. voltages *v1* and *v2* are both zero. To ensure that the circuit always ends up in the correct operation point a startup circuit is included. There, the resistor R_S is used to produce a current which is injected into node n1 if the voltage *v1* goes below one threshold voltage of an nMOS transistor. In the desired operation point the voltage *v1* is above the threshold and thus the startup circuit has no effect on the BGR circuit. Since the opamp is biased from the BGR, the startup circuit also ensures its bias current.

According to simulations [14], the circuit can be used with a supply voltage ranging from 0.95 V to 1.50 V and at temperatures ranging from -20 to $+100°$C.
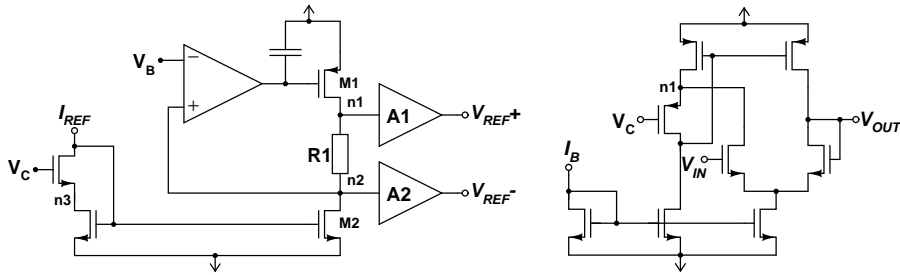
Figure 11.7 Differential reference voltage driver (left) and unity gain buffer A1 (right).

11.3.2   Reference Voltage Driver

The voltage provided by the bandgap reference is generated across a resistor, and thus it is not suitable for supplying a switched capacitor load without buffering. Typically, ADCs (pipelined or delta-sigma) based on the switched capacitor or switched opamp techniques utilize fully differential circuitry, which demands the reference voltage to be differential, i.e. a difference of two voltages set symmetrically between the supply rails. The circuitry in these ADCs operates in two phases, each lasting half of the clock cycle. Consequently, the capacitors have to be loaded to the reference voltages in half a clock period.

The proposed driver circuit is shown on the left in Figure 11.7. There, the reference current I_{REF} is supplied by the bandgap reference, which is the circuit shown in Figure 11.6 without the resistor R4. The current is mirrored with transistor M2 to go through a floating resistor R1, matched with the resistors in the bandgap reference circuit. As a result, the differential reference voltage appears across the resistor R1. The common mode voltage level is controlled by adjusting the current of M1 with an opamp and feedback loop. The bias voltage V_B is generated with a replica circuit. To improve the accuracy of the current mirroring, the voltage V_C is adjusted in such a way that the voltage of node n3 tracks node n2.

The generated reference voltages at nodes n1 and n2 are buffered with the unity gain buffers A1 and A2. The schematic of the buffer A1 is shown on the right in Figure 11.7. It is an amplifier, which consists of a differential pair and a low voltage current mirror load, connected in unity gain feedback. The cascode transistor is biased in such a manner that node n1 tracks the input voltage in order to minimize the systematic offset resulting from the amplifier imbalance. The buffer A2 is similar to A1, except that all nMOS transistors are replaced with pMOS devices and vice versa.

Chapter 12

Prototypes and Experimental Results

12.1 Measurement Setups and Methods

12.1.1 Measuring Dynamic Performance of S/H Circuits

The target application of an S/H circuit has a large impact on measuring the circuit. If the circuit is designed to be used with an ADC integrated on the same chip, no capability of driving an external 50-Ω load is needed. This, however, prevents straightforward full-speed measurements. On the other hand, in ADC applications, only the instantaneous value of the S/H circuit output at the end of the hold phase is of interest. Consequently, continuous time measurements may give results that are too pessimistic.

The easiest way to get rid of these problems is to characterize the S/H circuit together with the ADC. Then, however, it may be difficult to distinguish between the properties of the ADC and the S/H circuit. A 50-Ω driving capability can be obtained with an on-chip buffer, yet its implementation is often even more demanding than the design of the S/H circuit itself. So the designer can very easily end up in a situation where he or she is measuring the output buffer rather than the S/H circuit.

A widely-used way to characterize S/H circuits is the beat frequency test [37]. There, two S/H circuits are integrated on the same chip and one is used to measure the other. The measurement setup used to characterize the implemented circuits is shown in Figure 12.1. In this, the output of the first S/H circuit (the one on the left)
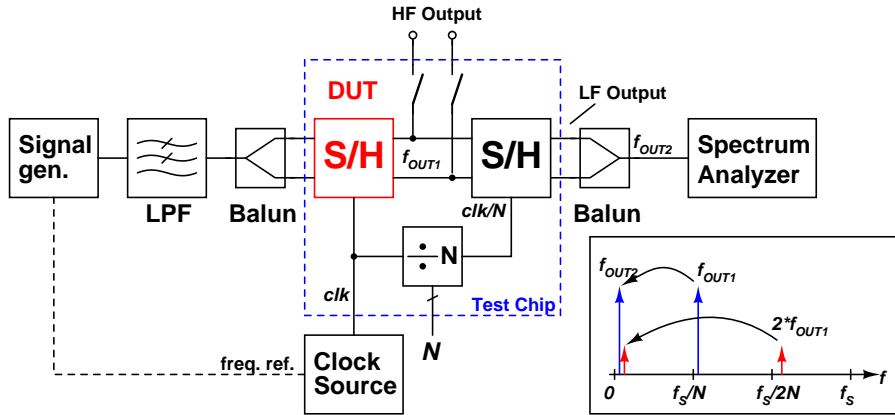
Figure 12.1 Measurement setup for beat frequency test.

is sub-sampled with the second circuit, whose clock signal is obtained by dividing the clock of the first circuit with an on-chip divider by some integer N. Now, if the input signal frequency is within a small offset, say Δf, of the clock frequency of the second circuit (f_S/N), the signal is aliased to a low frequency, which is equal to the frequency offset Δf. In a similar manner the possible harmonics in the first circuit output become aliased to the frequencies $2\Delta f$, $3\Delta f$, $4\Delta f$, and so on, as illustrated in the inset of Figure 12.1.

The beat frequency test provides a way to investigate the distortion characteristics of a S/H circuit at high signal frequencies without the need to bring high frequency signals out from the chip. The second S/H circuit does not even need to drive the 50-Ω input impedance of the measurement equipment or balun, since the low frequency signal can be handled with an active differential-to-single-ended converter constructed of discrete opamps.

Since the second S/H circuit samples the fully settled output of the first one, the measurement setup simulates the actual operation environment in front of an ADC on the same chip. Although the output of the second circuit is measured in continuous time, the error introduced is insignificant as a result of the fact that the difference between the concurrent sample values is very small. Also, the sinc attenuation resulting from the hold operation can be ignored at low frequencies.

12.1.2 ADC Measurements

The methods for measuring ADC performance and the related figures of merit are outlined in two IEEE standards. The more recent of them, IEEE-STD-1241 [220],

is specially targeted on ADCs, and thus it virtually replaces the earlier one [221] for waveform recorders.

12.1.2.1   Static Linearity

The quantization levels of an ADC can be measured using a servo loop. The linearity errors, DNL and INL, are calculated from the measured levels. Alternatively, a histogram-based method can be used to reduce the complexity of the measurement setup. In this, a signal with a known waveform is applied to the ADC input and the resultant output codes are collected into bins, each of which corresponds to one possible ADC output code. The number of samples that fall into a code bin represents the bin width, while the ideal bin widths can be derived with the knowledge of the signal waveform. The difference between these two is used to calculate the DNL and INL. This so-called code density test can be performed at full speed and with a sufficiently high signal frequency.

The histogram of a triangular or sawtooth wave is flat, making the calculations easy. Generating either type of signal with adequate purity, however, is difficult. Thus, a sinusoidal signal is often preferred, since it can readily be generated with a general signal source and proper filtering. Furthermore, the signal's purity can be checked with a spectrum analyzer.

Properly selecting the signal frequency with respect to the clock frequency is important in order to guarantee that the code density is not affected by unwanted correlation between the signal and the clock frequency. The required length of the data record is set by the noise and desired tolerance and confidence levels. A more detailed description of the code density test can be found from the standard [220] and an earlier publication [222]. The development history of the method can be followed with references [223, 224, 225, 226].

12.1.2.2   Signal to Noise and Distortion Ratio

The standard provides two methods for determining the SNDR, a frequency domain method based on DFT (discrete Fourier transform) and a time domain method using curve fitting. In the latter, the recorded sine wave is fitted to an ideal sine wave by minimizing the mean square error. The difference between the curves includes the ideal quantization error as well as the effect of static and dynamic ADC errors and noise. Thus, the SNDR can be calculated. This is the method used in measuring the prototypes described later in this chapter.

The alternative method extracts the error energy from the spectrum obtained with

the DFT. The signal energy is in one frequency bin, while the error energy is distributed to the others, with the exception of the zero frequency bin, which also contains the DC term. The method requires the signal energy to be contained exactly in one bin, which is obtained when the record has an integer number of signal cycles. The sine fit method does not have this restriction.

The SFDR and THD can be determined from the DFT spectrum.

12.2 S/H Circuit Using Double-Sampling

The goal of this design is to develop a high-speed CMOS S/H circuit for time-interleaved ADCs. The target specifications were set as follows: 10-bit resolution, a sampling rate higher than 100 MS/s, 2-Vpp differential signal swing from a 3.0-volt supply, and reasonable power consumption. This prototype has been published in [12, 10, 11].

12.2.1 Architecture

The architecture of the prototype is shown in Figure 12.2. It is a fully differential version of the double-sampled S/H circuit shown in Figure 9.3. The differential structure is almost a necessity in the mixed signal environment of ADCs, where the amount of substrate noise and other disturbances is considerable.

The signal common mode level at the input and the output of the circuit need not be the same; neither do the common mode level at the input and the output of the opamp. This can be utilized to adjust the level of the continuous time input signal in the region where the distortion caused by the signal-dependent switch on-resistance is minimized. In the case of nMOS switches, the input signal level should be as small as possible. There are, however, two reasons which set a lower limit for the input signal common mode level. First, the negative peak voltage may not go much below V_{SS}, in order to prevent the pn-junctions in the drain and the source of the MOS switch from becoming forward-biased. On the other hand, if the S/H circuit is driven without DC decoupling the driver circuit probably cannot provide a signal swing that ranges down to V_{SS}. The signal levels used in this design are shown in Figure 12.3.

The offset between the opamp input and output common mode level in hold mode is the same as the difference between the input signal CM voltage and the sampling ground. The maximum tolerable offset is heavily dependent on the type of opamp. From the sampling switch point of view, it is preferable to make the sampling ground voltage as low as possible in order to reduce switch size. For maximum signal swing
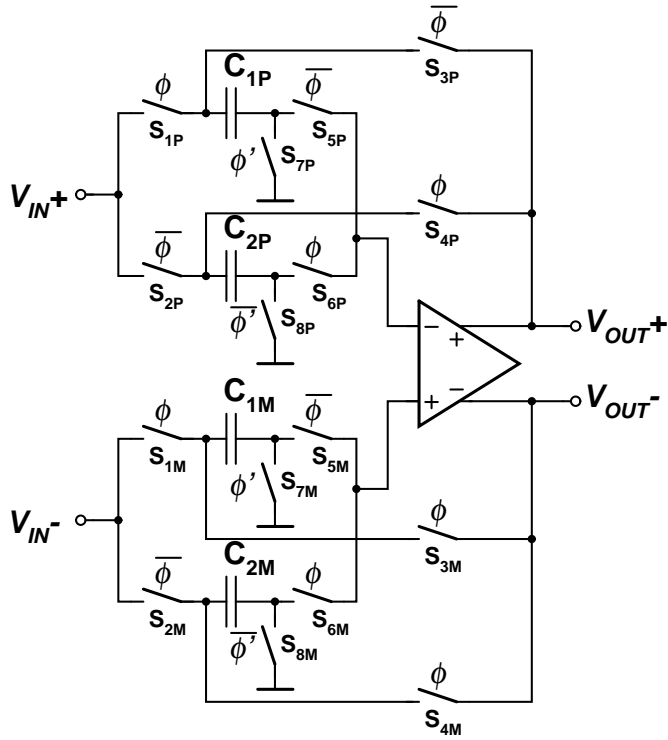
Figure 12.2 Fully differential double-sampled S/H circuit.
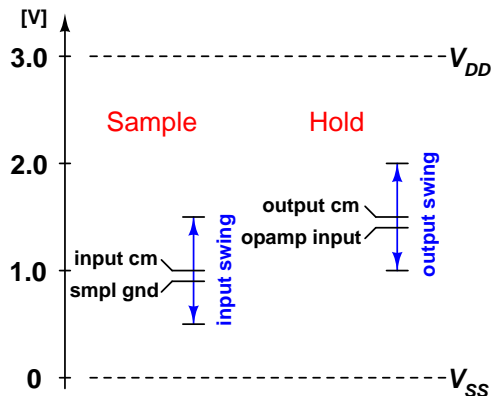

Figure 12.3 Signal ranges and common mode levels in the circuit.

and minimum distortion the output common mode level is set half-way between the supply voltages. The circuit employs the bottom plate sampling technique to avoid a signal-dependent charge injection from the MOS switches.

12.2.2 Switches

The switches throughout the design are implemented with nMOS transistors. The distortion resulting from the signal-dependent switch time constant is suppressed below the target level by controlling the switches with a voltage higher than the 3-volt supply. This voltage is not generated on the chip; instead, an external voltage source is used.

Although bottom plate sampling minimizes the signal-dependent charge injection, the constant common mode voltage step resulting from the injected charge and clock feed through can still be a problem, since it may cause the common mode level in the opamp input to exceed the valid range. By making the switches S_1 and S_8 equal in size, the problem can be minimized, thanks to the fact that the switches operate in opposite clock phases, and thus their charge injections cancel each other.

12.2.3 Clock Generator

Avoiding systematic timing skew in the double-sampled S/H circuit is essential. Thus, so as to guarantee an exact $180°$ phase difference in the half-rate clocks, the input for the clock generator is derived from the incoming full-rate clock using a synchronous divide-by-two circuit, which is built with a differential D-flipflop [227]. The clock generator relies on a standard structure, based on cross-coupled OR-gates, in producing the non-overlapping clock phases. The last clock buffer stages, which provide the switch control voltage, use the high supply voltage.

12.2.4 Opamp

The opamp architecture, which has already been studied in Section 7.2.3, is based on a cascode output stage and low-gain first stage. Since it is fully differential, it requires a common mode feedback circuit. Due to the double-sampling, the common mode feedback has to be active in both the clock phases. Such a feature is easily implemented with a continuous-time CMFB circuit. In this design, however, two parallel switched capacitor CMFB circuits are operated in opposite clock phases. The circuit is shown in Figure 12.4.
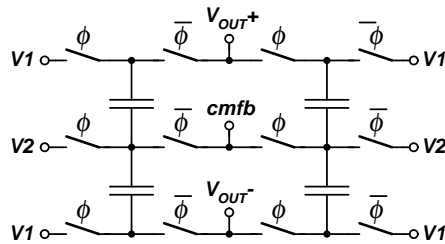
Figure 12.4 CMFB circuit for double-sampling circuit.

The simulated opamp frequency response shows a 450-MHz GBW and 62-degree phase margin at the unity gain frequency and about a 70-degree one at the frequency of closed loop gain in the target feedback configuration. The DC gain simulated with the nominal transistor parameters is 62 dB and the settling time to 10-bit accuracy 4.5 ns.

12.2.5   Experimental Results

The circuit was fabricated with a 0.5-μm double-poly triple-metal CMOS process. A photograph of the prototype chip, containing two S/H circuits and a programmable divider, is shown in Figure 12.5.

The circuit is characterized with the beat frequency test with several sub-sampling ratios. A spectrum of a 73.3-MHz, 1.75-Vpp signal, which is sampled at 220 MS/s, is shown in Figure 12.6. The SFDR is limited by the third harmonic, which in this case is at the 65.6-dBc level.

In addition to the signal and its harmonics, there is an extra spurious frequency at the 25-kHz offset from the signal peak. It is not generated by the S/H circuit under test. The same spur (and also a number of its multiples) is seen in the spectrum of the signal generator, which is used as the clock source. Thus, it probably originates from the leakage of the PLL reference in the signal generator.

The SFDR is measured as a function of the signal amplitude at 130 and 220-MS/s sampling rates. The results are shown in Figures 12.7 and 12.8 respectively. There are two curves in both the figures, a solid curve for a 200-kHz input signal and a dashed curve for an input signal which is within a small frequency offset of one third of the clock frequency. The results show that the circuit operates well at both clock rates. As expected, the SFDR decreases as the signal amplitude is increased. At 220 MS/s 10-bit resolution is achieved with a 1.8-Vpp signal from DC up to one third of the clock frequency.
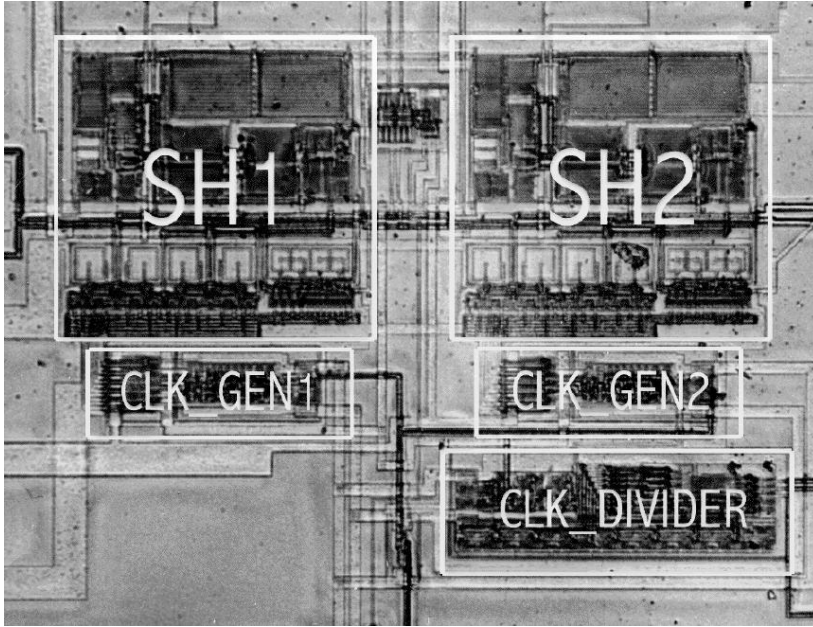
Figure 12.5 A photograph of the double-sampling S/H prototype.
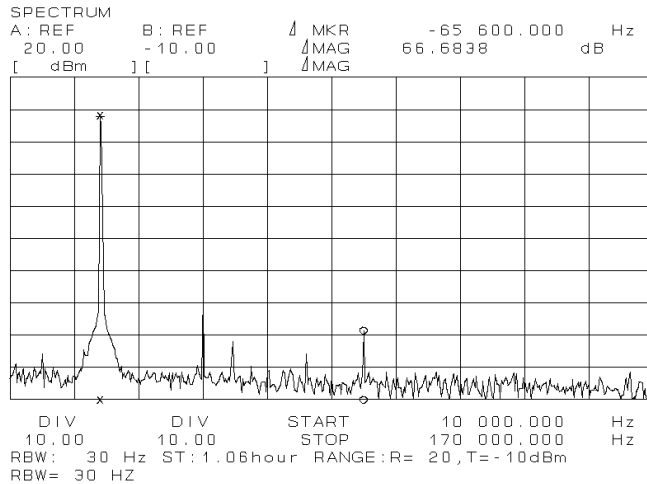
Figure 12.6 Measured spectrum of a 73.3-MHz @ 1.75-Vpp signal sampled at 220 MS/s. The aliased signal is seen at the 33.3-kHz frequency and its harmonics at multiples of that frequency. The spur at a 25-kHz offset from the signal peak does not originate in the S/H circuit.

Figure 12.7 SFDR as a function of the signal amplitude at a 130-MS/s sampling rate.

Figure 12.8 SFDR as a function of signal amplitude at a 220-MS/s sampling rate.

Figure 12.9 SFDR of a 73.3-MHz signal sampled at a 220-MS/s rate as a function of the switch control voltage.

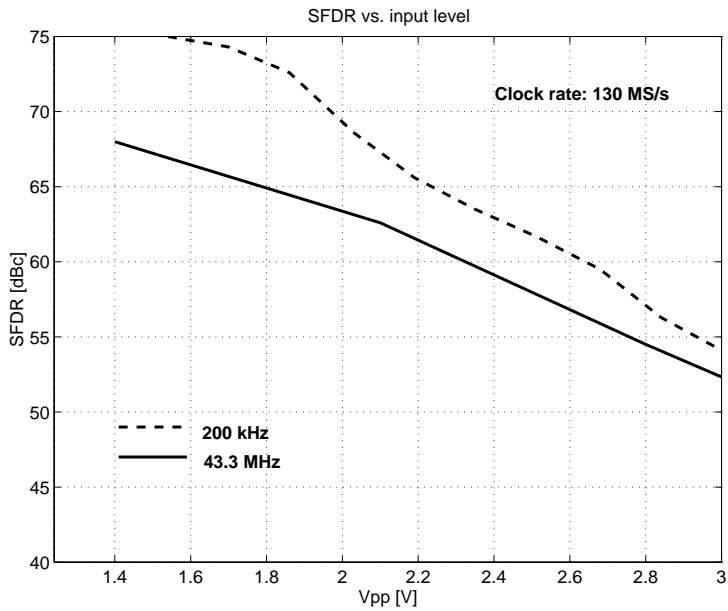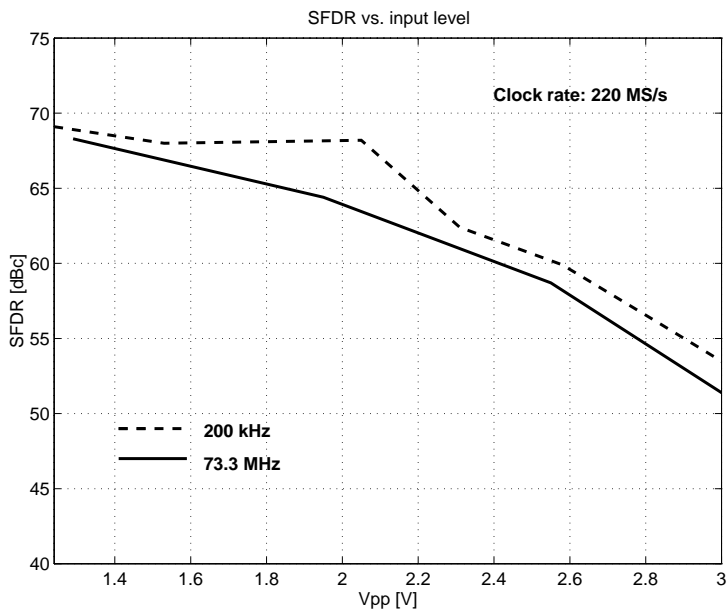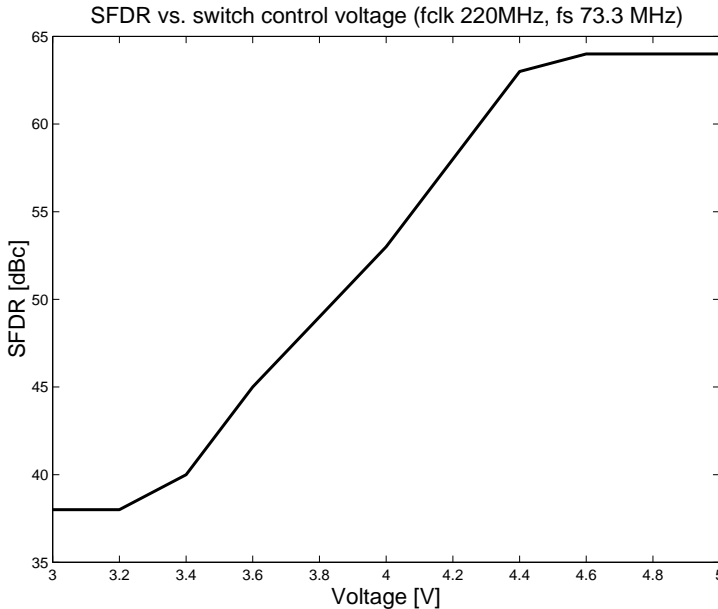Limitations in the measurement equipment prevented the testing of the circuit at clock rates higher than 220 MS/s. Neither was it possible to measure the circuit with a $f_S/2$ input signal at 220 MS/s because of the lack of a proper filter to remove the harmonics from the test signal. However, at 130 MS/s, the SFDR was even slightly better with the $f_S/2$ than with the $f_S/3$ input signal, which is probably due to the fact that with the $f_S/2$ input frequency the voltage on the sampling capacitors is almost unchanged between the samples.

An interesting study is the effect of the switch control voltage on the SFDR. The measurement results obtained with a 73.3-MHz signal at a 220-MS/s sampling rate are shown in Figure 12.9. The SFDR dependence on the control voltage is almost linear from 3.2 V to 4.4 V. Increasing the voltage above 4.5 V does not give any improvement, which indicates that the distortion from the other sources starts to dominate at that level. The 4.5-V control voltage results in a 4-V maximum switch transistor gate-drain voltage, which is a couple of hundred millivolts larger than the maximum long-term reliable value.

The measurements with odd sub-sampling ratios revealed that there is a spurious tone at $f_S/2 - f$. It most probably originates from the timing skew between the parallel circuits. In the worst case, when the input signal frequency is from a small offset of

Table 12.1 Measured performance of the double-sampling S/H.

Sampling rate	220 MS/s
SFDR	65 dBc
Differential input swing	1.8 Vpp
Supply voltage	3.0 V
Power consumption	25 mW
Active area	0.06 mm^2
Technology	0.5-μm CMOS

the Nyquist frequency, the level of the spur is -61 dBc, which corresponds to a 2.6-ps timing skew.

The power consumption of the circuit without the clock generator was measured as 25 mW at a 220-MS/s sampling rate with a $f_S/3$, 1.75-Vpp input signal. The performance of the circuit is summarized in Table 12.1.

12.3 Timing Skew-Insensitive Double-Sampling S/H

In order to avoid the timing skew problem another version of the S/H circuit, employing the skew-insensitive sampling proposed in Section 9.3, was designed and tested. This prototype has been reported in [9].

12.3.1 Architecture

The architecture of the circuit is a fully differential version of the timing skew-insensitive circuit proposed in Section 9.3. For the sake of convenience it is shown again in Figure 12.10. The building blocks, except the clock generator, (opamp, switches, CMFB, etc.) are from the earlier prototype.

12.3.2 Clock Generator

The new clock generator is shown in Figure 12.11. The circuit generating the non-overlapping signals is basically the same as used in the first S/H circuit. The short pulses for the common sampling switch are constructed with a circuit consisting of an inverter, a delay element, and a NAND gate. The D-flipflop generates the complementary half-speed clock signals.

To reduce the jitter in the sampling clock (ϕ_S), the buffer chain can be made shorter by connecting the clock input of the D-flipflop directly to the incoming clock.

Figure 12.10 Timing skew-insensitive double-sampling S/H circuit.

Figure 12.11 Clock generator for the skew-insensitive S/H circuit.
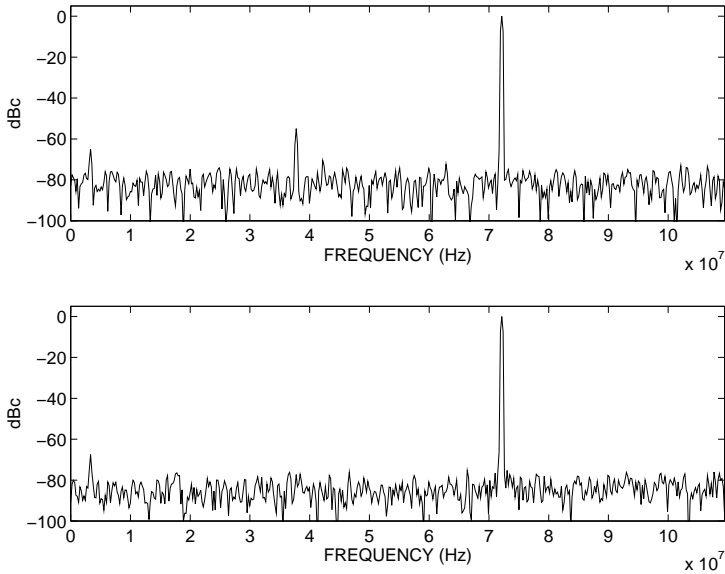
Figure 12.12 The simulated effect of a 10-ps timing skew on the spectrum of the first (upper plot) and the second (lower plot) prototype circuits.

12.3.3 Simulations

The effect of timing skew on the first and second S/H circuits is compared by adding an intentional 10-ps timing skew in the clock signals. As expected, this has no effect on the skew-insensitive circuit. The FFT spectra for both circuits calculated from transient simulations are shown in Figure 12.12. The clock frequency in the simulations is 220 MS/s and the signal frequency one third of that. An error image with a 53-dBc magnitude is seen at the 38-MHz frequency in the output of the first circuit. This is exactly as predicted by the theory. There is no sign of this image in the spectrum of the skew-insensitive circuit. Except for the image, the two spectra are almost identical, which indicates that the new switching scheme does not degrade the other properties of the circuit.

12.3.4 Experimental Results

The test chip was fabricated with the same 0.5-μm CMOS process as the first chip. A photograph of the chip is shown in Figure 12.13. It turned out that an unfortunate mistake was made in the beat frequency test setup design; the same programmable divider that was used with the first prototype was applied without any modifications. The new clock generator, however, now included a frequency division by two and, as
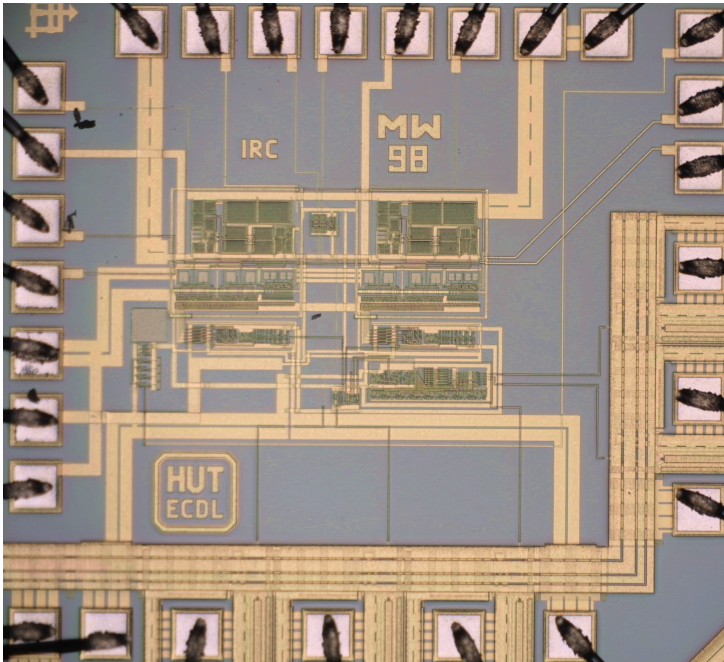
Figure 12.13 Photograph of the skew-insensitive S/H prototype.

a result, the second S/H circuit on the chip could be clocked only with the rates $f_s/4$, $f_S/6$, $f_S/8$, etc. The odd sub-sampling ratios would have been needed to investigate the spectrum image resulting from the timing skew. This is not possible with the even ratios, since then the image aliases at the top of the fundamental signal.

Although the circuit was later characterized as a part of a time-interleaved ADC (see next section), the absence of a skew-originated spurious image could not be totally verified. The measurement results showed an image, the magnitude of which was highly dependent on the ADC bias settings and thus it was likely originating from sources not related to timing skew. If there was an image produced by the skew, it was masked by this image.

Although the elimination of timing skew could not be verified with this prototype, the circuit's performance was measured with the available sub-sampling ratios. The results show that the performance is almost identical to the first prototype. This proves that the timing skew-insensitive switching does not degrade other circuit characteristics. A spectrum where a 1.8-Vpp, $f_S/4$ signal is sampled at 220 MS/s is shown in Figure 12.14. Again, the spurs at the 25-kHz offset from the fundamental are due to a poor quality clock source.
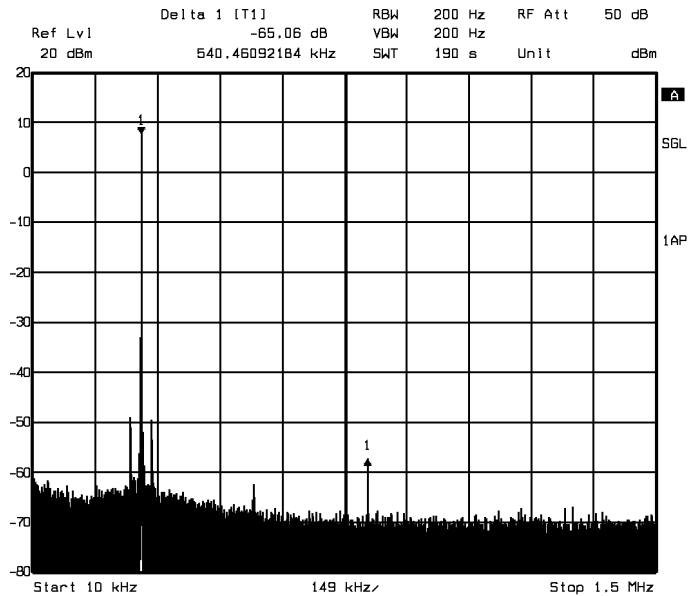
Figure 12.14 Measured spectrum of the timing skew-insensitive S/H circuit. The spectrum is measured at a 220-MS/s sampling rate with a 1.8-Vpp input signal having a frequency which is a small offset from $f_S/4$. The spurs close to the carrier are part of the spectrum of the used clock source.

12.4   10-Bit, 200-MS/s Parallel Pipeline ADC

The most promising topology for a high-resolution high-speed CMOS ADCs is pipeline architecture. Parallel pipeline ADCs with several time-interleaved component ADCs have been introduced to attain very high sampling rates with acceptable power consumption [93, 92]. A resolution of ten bits and conversion rates up to 100 MS/s have been reported [98]. Their power dissipation, however, has risen very high, especially with high sampling rates. By employing double-sampling and parallelism with time-interleaved pipeline ADCs, a very competitive power and area consumption can be obtained [97].

The goal set for the prototype presented in this section was to demonstrate 10-bit resolution at a 200-MS/s sampling rate using a 0.5-μm CMOS technology. The design has been published in [5] and [4].

As discussed in Section 4.5.6, in a single-channel pipelined ADC the one effective bit per stage architecture gives a maximum conversion rate and minimum power consumption when capacitor scaling is not used. It is obvious that increasing the number of parallel channels raises the conversion rate and lowers the slew rate and

bandwidth requirements of the amplifier in a switched-capacitor (SC) gain stage. The current consumption of an operational amplifier is a nonlinear function of the bandwidth, which suggests that for a given technology there exists an optimum degree of parallelism with respect to the sampling rate and power dissipation. For the target 200 MS/s sampling rate the optimum number of channels was found to be four [4].

12.4.1   ADC Architecture

The well-known problems in time-interleaved ADCs arise from mismatch between the parallel channels. These errors are offset, gain mismatch, and skew in the clock signals. Offset is seen as tones at multiples f_S/M, where M is the number of parallel channels and f_S the sampling rate. Both gain mismatch and timing skew generate spectral images of the signal around the same frequencies. It is not possible to achieve a 10-bit resolution with the 200 MS/s sampling rate without performing special actions to eliminate these errors.

The most straightforward way to avoid timing skew, also used in this design, is to employ a front-end sample-and-hold (S/H) circuit. Digital calibration is chosen for the purpose of eliminating the offset, which mainly originates from the offset voltages of the operational amplifiers. The gain error, arising predominantly from capacitor mismatch, is left uncalibrated since it can be adequately suppressed by a careful layout design for the 10-bit accuracy requirement.

A block diagram of the 10-bit 200 MS/s pipeline ADC is shown in Figure 12.15 and its clock signals in Figure 12.16. The core of the converter consists of two parallel double-sampling pipeline ADCs. The differential analog input is time-interleaved to the four component ADCs in the order indicated in Figure 12.16 by a double-sampling S/H circuit. The digital outputs of the stages of the parallel ADCs are corrected and multiplexed to two 100-MHz time-interleaved outputs, which are offset compensated.

12.4.2   Front-End S/H Circuit

The front-end S/H circuit is based on the skew-insensitive S/H prototype described in the previous section. The only difference is in the switches; to improve the linearity and the reliability of the circuit, a separate high supply voltage for the switches is not used any more. Instead, the input switches, which suffer the most from the insufficient gate overdrive, are realized using bootstrapped switches, shown in Figure 12.17 [136]. The same circuit is also used as the input switch in the first pipeline stages, which have to track the input signal in a quarter of the clock period.
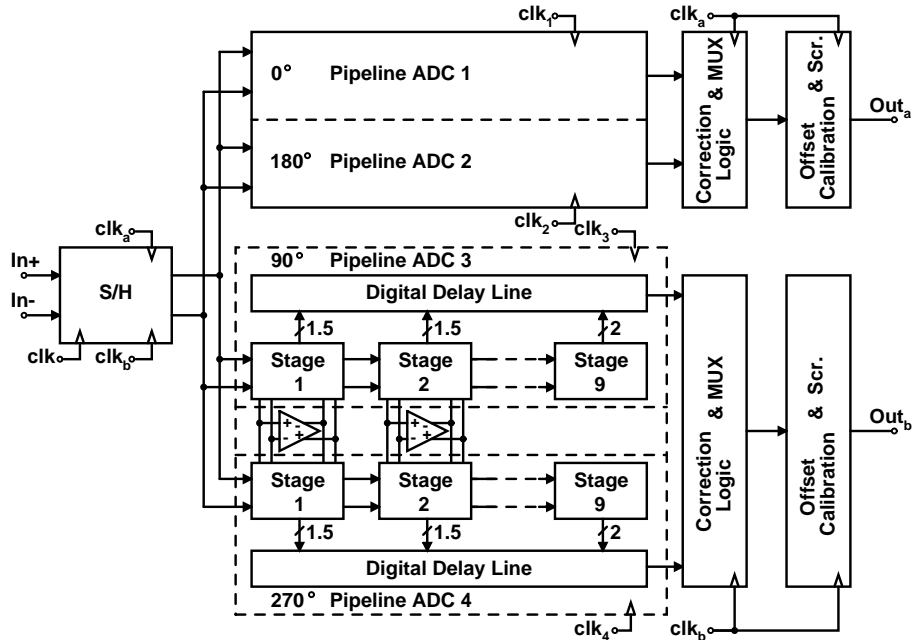
Figure 12.15 Block diagram of the 4-channel parallel pipeline ADC.
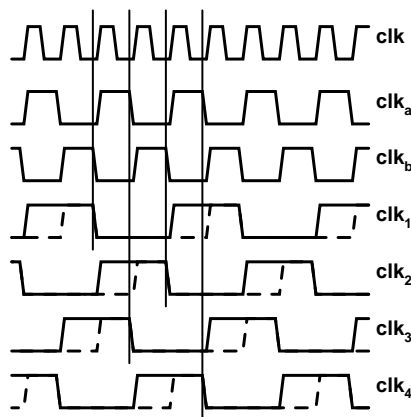
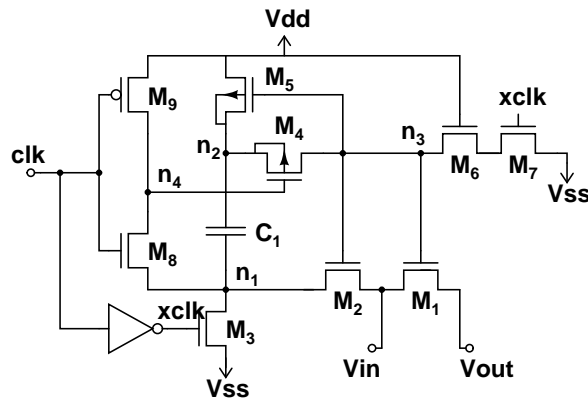**Figure 12.16** Clock signals for the parallel pipeline ADC.

Figure 12.17 Bootstrapped switch.

Since the feedback factor in the flip-around S/H architecture is close to one, the S/H circuit achieves twice the speed of a pipeline stage using the 1.5-bit architecture, where the feedback factor is ideally 0.5. Thus, the S/H circuit can drive two parallel ADC channels without becoming a bottleneck for the conversion rate. Utilizing double-sampling in both circuit blocks allows the number of channels to be increased to four.

12.4.3 Component ADCs

The four pipeline component ADCs employ the 1.5 bit/stage topology with RSD error correction. The eight pipeline stages are followed by a two-bit flash ADC, as indicated in Figure 12.15. The property of the successive pipeline stages working in opposite clock phases is exploited by sharing the operational amplifiers between two parallel component ADCs.

In the double-sampling MDAC, shown in Figure 12.18, the capacitor arrays, operating in a 180° phase shift, have their own three-level sub-ADCs, which minimizes the delay of the switch control signals but doubles the number of comparators needed. However, the dynamic comparators used in the sub-ADC have a very small power dissipation and area. In the MDAC the switches whose timing is critical and the switches connected to the input of the amplifier are nMOS switches. In all other switches a more linear response in the signal voltage range and minimization of clock feedthrough errors are achieved by utilizing CMOS switches. In the first pipeline stage the input switches tracking the output of the S/H circuit are realized with a bootstrapped MOS switch.

Figure 12.19 shows the opamp, which uses the folded cascode architecture with
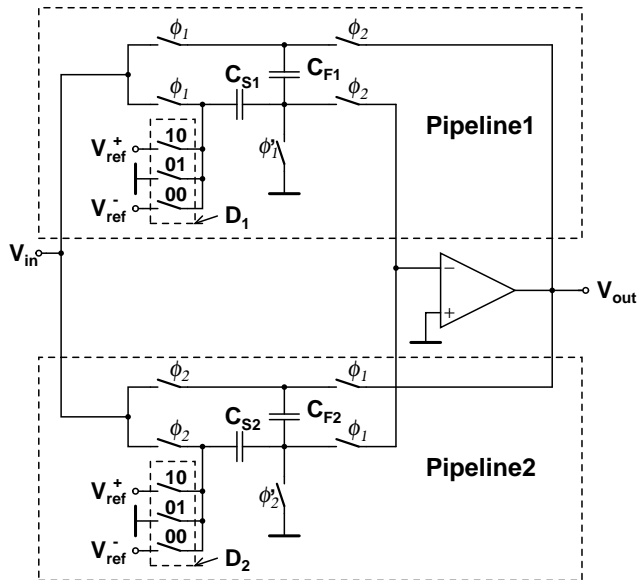
Figure 12.18 Two MDACs share a common opamp.

regulated cascode devices. Regulation amplifiers based on a common gate input structure make possible the biasing of the cascode nodes at one V_{dsat}, thus providing high output signal swing.

Measurements of the first version of the prototype revealed that many of the comparators had offsets exceeding the correction range of the RSD logic, which in this 1.5 bit/stage architecture is as high as ± 200 mV. This indicated that there was something wrong with the dynamic comparator circuit, which is based on input devices biased in the triode region [73]. Simulations showed that the structure is rather sensitive to mismatch in the devices forming the regenerative latch. Furthermore, the sensitivity is strongly dependent on the signal common mode voltage.

To get rid of the errors caused by the offsets a new comparator, shown in Figure 12.20, was designed and used in the second version of the prototype. It uses two differential pairs with pulsed current sources to form currents proportional to the difference between the differential reference voltage and differential signal voltage. The currents are summed and fed into a latch. The circuit is not sensitive to common mode voltages, even when they are not the same in the signal and in the reference, and the offset is primarily determined by the mismatch of the input devices. This is verified by measurements of separately processed test structures [228].
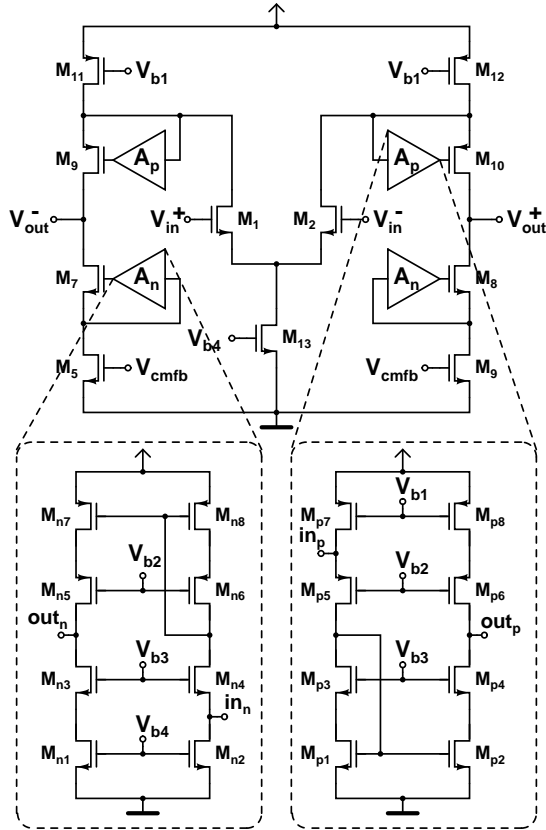
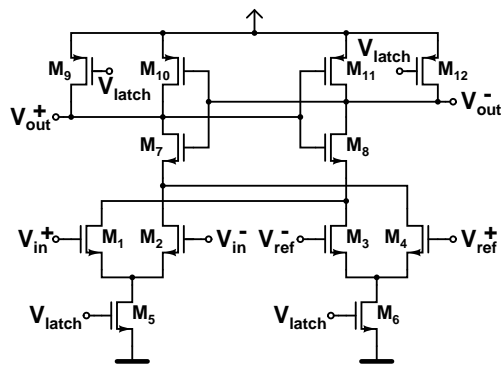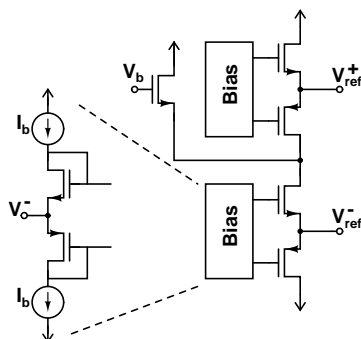**Figure 12.19** Folded cascode OTA with high-swing regulated cascode devices.


Figure 12.20 Differential pair dynamic comparator.

Figure 12.21 Simplified schematic of the reference voltage driver.

12.4.4 Reference Voltage Driver

The large number of pipeline stages using common reference voltages increases the capacitive load in the reference nodes to several picofarads. To guarantee that the reference does not limit the settling speed, its output impedance has to be in the order of a couple of dozens of ohms. This means that a resistor string implementation of the reference would have a very large quiescent current or, alternatively, a large external capacitor has to be used. More reasonable power consumption without an external capacitor is obtained with the circuit proposed in Figure 12.21. There, the low impedance voltage outputs are provided by class AB buffers constructed of complementary transistors. The buffers for the positive and the negative reference are cascaded in order to minimize steady state current consumption.

12.4.5 Digital Offset Calibration

The main source of offset is the input offset voltage of the operational amplifiers utilized in the pipeline stages. Because of the double-sampling there is no idle time which could be used to auto-zero the amplifier offset and thus the problem has to be handled in some other way. The methods that can be used to suppress the offset in parallel ADCs include digital [93] and analog [99] calibration and digital post filtering in the case of a two channel ADC [95]. In analog calibration the offset is measured from the digital output and, using a D/A converter, a canceling signal is injected into the channel input. Digital calibration does the canceling in the digital domain, simply by subtracting the measured offset from each sample. The advantage of the analog method is that it does not reduce the signal range. However, the robustness achieved with digital calibration usually makes this method preferable, despite the small reduction in the signal range.
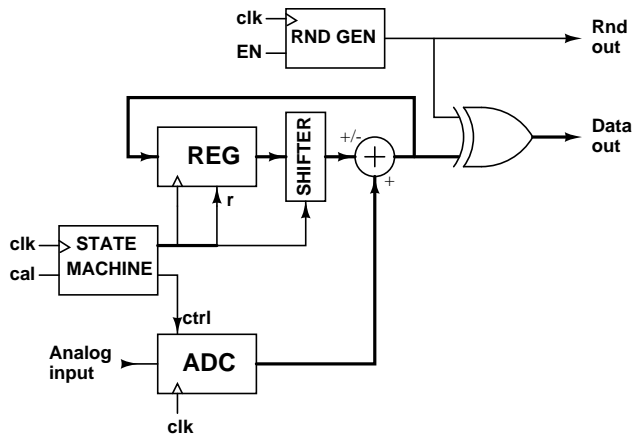
Figure 12.22 Offset calibration and scrambling of the output.

In this design the digital offset calibration is applied in the multiplexed half-rate output of the double-sampled channel pair. It is assumed that there is no significant offset, and thus no need for calibration, between the signal paths inside the double-sampled pipeline. This is due to the fact that the opamps, which are the main source of the offset, are the same for both the signal channels. During the offset measurement the normal operation of the converter has to be suspended. The calibration can, however, usually be performed at the power-up or during some idle periods.

The calibration circuit consists of an adder, a register for storing the measured offset, and a state machine that provides the control signals for the logic and the ADC. During the calibration the ADC input is shorted to ground and the offset is obtained by averaging the output signal over 16 clock cycles. The averaging is realized simply by adding together 16 consecutive output codes and performing a bit shift of four for the result. The calibration is activated with an external one-bit control signal.

The most significant ADC output bits have a strong correlation to the analog input signal. This is utilized to investigate the signal feedthrough from the output to the input by adding the possibility of scrambling the outgoing digital words with a pseudo-random bit-stream [229]. The scrambling is realized by putting XOR gates before each output buffer and applying the random bit to their other input. For unscrambling, the random bits are taken out through an extra package pin. A simplified block diagram of the calibration and scrambling circuit is depicted in Figure 12.22.
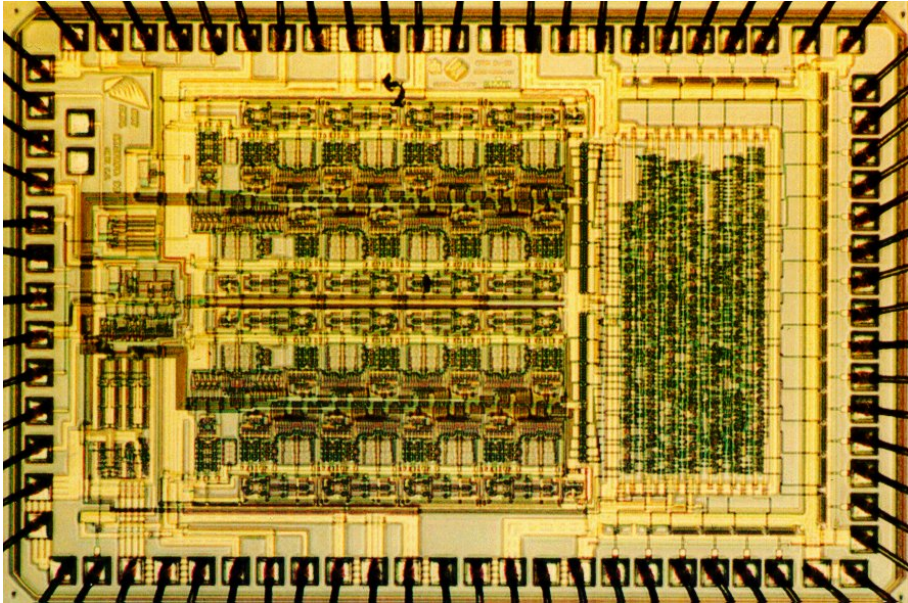
Figure 12.23 Parallel pipeline ADC chip micrograph.

12.4.6 Experimental Results

The prototype circuit is fabricated using a 0.5-μm triple-metal double-poly CMOS process. The total area of the chip is 7.4 mm^2 and its die photograph is shown in Figure 12.23. The circuit is measured with a 3.0-V supply with a differential input swing of 1.6 V$_{pp}$.

The static linearity curves obtained with the code density test are presented in Figure 12.24, which shows the DNL as being within ± 0.8 LSB and the INL within ± 0.9 LSB. A spectrum obtained with a 199.975-MHz beat frequency at a 200-MHz clock rate is shown in Figure 12.25. In Figure 12.26 a 71.3-MHz full-scale signal is sampled at the full clock rate, resulting in a spurious-free dynamic range (SFDR) of 55 dB. From the spectrum it can be seen that the offset tone at 50 MHz is more than 56 dBc below the signal level and the gain mismatch tones around $f_S/4$ and $f_S/2$ remain below the noise level. The mismatch tone in the vicinity of half the sampling frequency usually rises to limit the SFDR at high signal frequencies. The total harmonic distortion (THD) as a function of signal frequency is plotted in Figure 12.27. THD starts from 55 dB, becoming about 46 dB around the Nyquist frequency, then rising again to the 55-dB level near the sampling frequency, implying that the performance of the ADC is limited by the pipeline component ADCs rather than by the S/H circuit.

DIFFERENTIAL NONLINEARITY

INTEGRAL NONLINEARITY
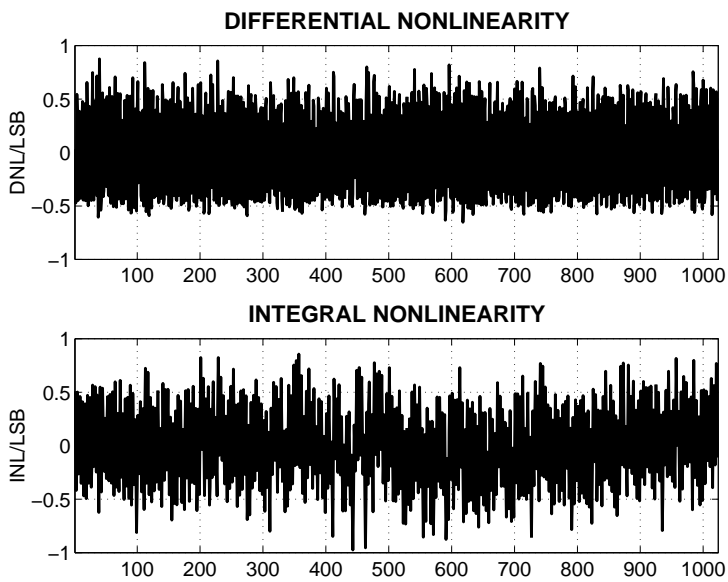
Figure 12.24 Measured DNL and INL.

POWER SPECTRUM

Figure 12.25 Spectrum obtained with a 200-MHz beat frequency test.

Figure 12.26 Measured spectrum where a 71.3-MHz signal is sampled at a 200-MHz clock rate.
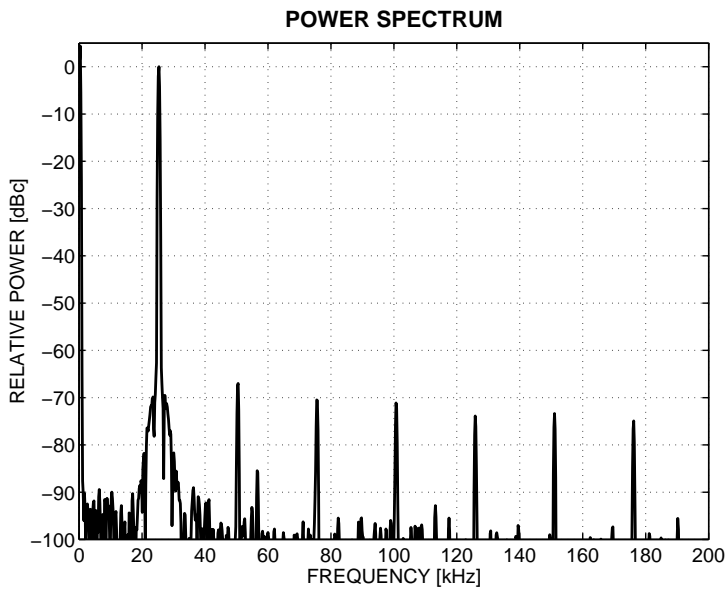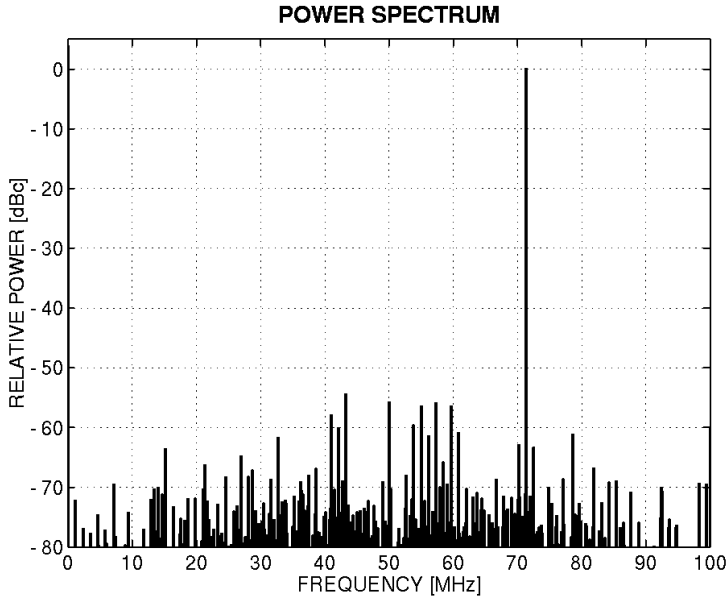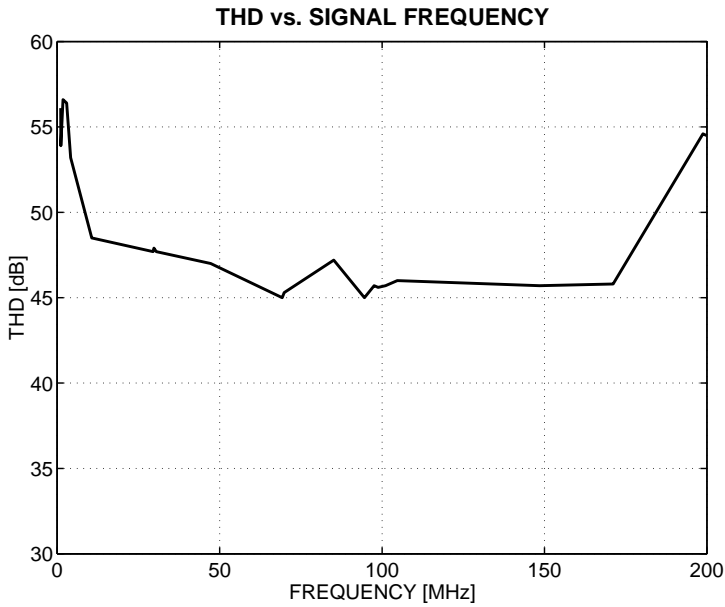

Figure 12.27 THD as a function of signal frequency.

Table 12.2 Summarized performance of the ADC.

Technology	0.5-μm CMOS
Resolution	10 bits
Sample Rate	200 MS/s
Supply Voltage	3.0 V
Area	7.4 mm^2
DNL	$\pm$0.8 LSB
INL	$\pm$0.9 LSB
THD	46 dB
Power Dissipation of Core ADC	280 mW
Power Dissipation including Output Buffers	405 mW

The mismatch tone probably results from improperly-operating common mode feedback, which creates asymmetry between the two double-sampled pipeline ADCs. This also limits the SNDR at high signal frequencies to 43 dB. However, excluding the mismatch tone, the spectral performance indicates that the ADC can sample narrow IF bands around 200 MHz with a THD of 55 dB. The measured power consumption without and with the digital output drivers are 280 mW and 405 mW at a 3.0-V supply, respectively. Table 12.2 summarizes the overall performance.

12.5   13-Bit Self-Calibrated IF-Sampling Pipelined ADC

The increasing bit rates in both wired and wireless telecommunication systems are made possible by utilizing wider signal bandwidths. Simultaneously, there is a desire to realize an increasing portion of the receiver functions in the digital domain. These trends lead to more and more demanding specifications for A/D converters. Typically, a resolution of from 12 to 15 bits is needed at input signal frequencies of dozens of megahertz and with a sampling rate of 50 MHz or higher. An important application area is 3rd generation cellular base stations, where the current trend is to move the analog-digital boundary to the intermediate frequency (IF), which makes the design of the A/D front-end even more challenging.

The requirement to sample an IF signal sets stringent specifications for the analog front-end of the ADC. Usually it is realized with a sample-and-hold (S/H) circuit which has to be able to track the high frequency input signal. One of the major challenges in implementing the S/H is the high frequency linearity of the sampling circuit, which is mainly determined by the properties of the sampling switch. In IF sampling the signal down conversion is performed by the sampling operation and thus the jitter of the LO signal, which is used as the sampling clock, must be very low.
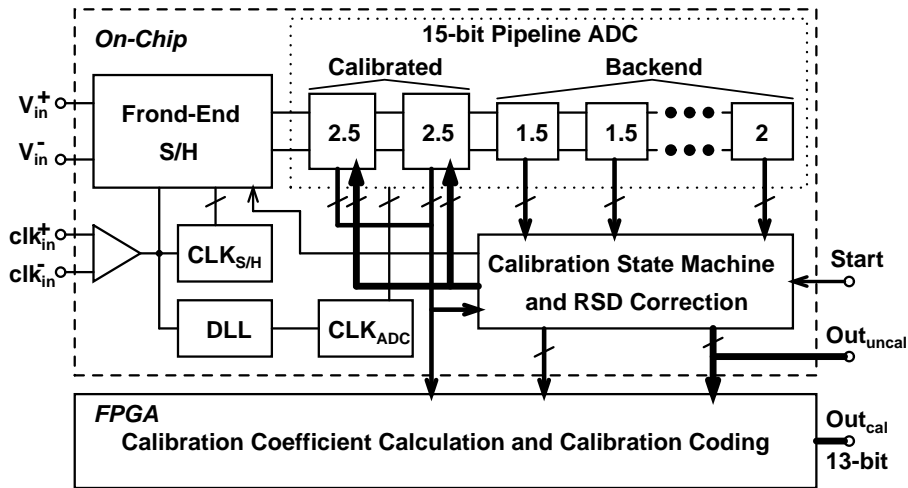
Figure 12.28 Block diagram of the self-calibrated IF-sampling ADC.

Typically, the matching properties of the circuit elements set the maximum attainable ADC resolution (with a reasonable yield) somewhere between 10 and 12 bits. To achieve a higher resolution with a Nyquist rate ADC, some kind of calibration or trimming has to be applied. Recently digital self-calibration techniques, which are made feasible by the possibility of including more and more digital circuitry in the ADC, have gained in popularity.

Originally published in [1], the prototype presented in this section is a 13-bit pipeline ADC incorporating a digital self-calibration algorithm. The ADC has a front-end S/H circuit designed to sample signals from a 200-MHz IF. In order to cope with thermal noise, the signal range is set as high as 3.8 V differential. Because the circuit operates on a 2.9-V supply, the large signal range has a major impact on the circuit structures used in the opamps, comparators, and switches.

12.5.1   Architecture

The block diagram of the ADC is depicted in Figure 12.28. It consists of an IF-sampling front-end sample-and-hold circuit, a self-calibrated 13+2-bit pipeline ADC, a delay locked loop (DLL) for synchronizing the sampling of the first pipeline stages and the S/H output, and an on-chip calibration state machine for controlling the S/H circuit and the first four stages of the pipeline during the calibration cycle. The two extra bits are used internally for the calibration. Redundant sign digit (RSD) coding is exploited to relax the comparator offset specifications. Calculation of the calibration
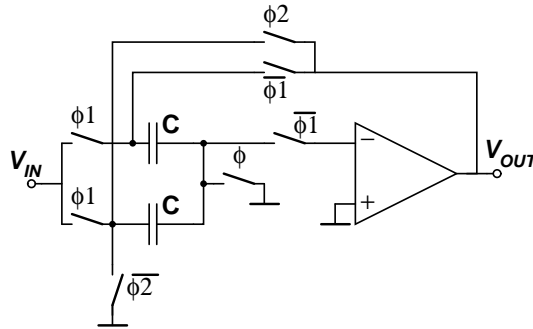
Figure 12.29 Front-end S/H circuit.

coefficients and calibration coding is realized with an external FPGA circuit.

12.5.2 Front-end S/H Circuit

The performance of the ADC at high signal frequencies is predominantly set by the front-end S/H circuit. Since it is in front of the signal chain, its thermal noise and distortion are not attenuated by any preceding gain stages and thus it has to fulfill the full resolution requirement.

The S/H circuit, shown in Figure 12.29, is an SC amplifier with a programmable gain of 1 or 2. In the unity gain mode the circuit acts as a flip-around S/H circuit, where the input voltage is sampled into the capacitors during the sampling phase ($\phi, \phi1 = 1$) and in the hold phase ($\phi, \phi1 = 0$; $\phi2 = 1$) the capacitors are connected to a feedback loop around the opamp. In the gain-of-two mode the sampling phase is unchanged, but in the hold phase one of the capacitors is connected to the opamp output and the other to the signal ground ($\phi, \phi1, \phi2 = 0$). Now a charge transfer occurs from the grounded capacitor to the feedback capacitor and, as a result, the sampled voltage is amplified by the ratio of the total capacitance to the feedback capacitance. The circuit utilizes the bottom-plate sampling technique, where the sampling switch (controlled with ϕ) is opened slightly before the input switches (controlled with $\phi1$) so as to avoid signal-dependent charge injection from the input switches.

The two modes set different requirements for the opamp; the unity gain mode calls for unity gain stability, while the gain-of-two mode doubles the bandwidth requirement as a result of a smaller feedback factor. Thus, since the opamp has to fulfill both of the requirements, its specifications are more stringent than in either mode alone. The input-referred thermal noise in the sampling phase is the same in both modes, but the peak signal-to-noise ratio is 6 dB lower in the gain-of-two mode as a result of the
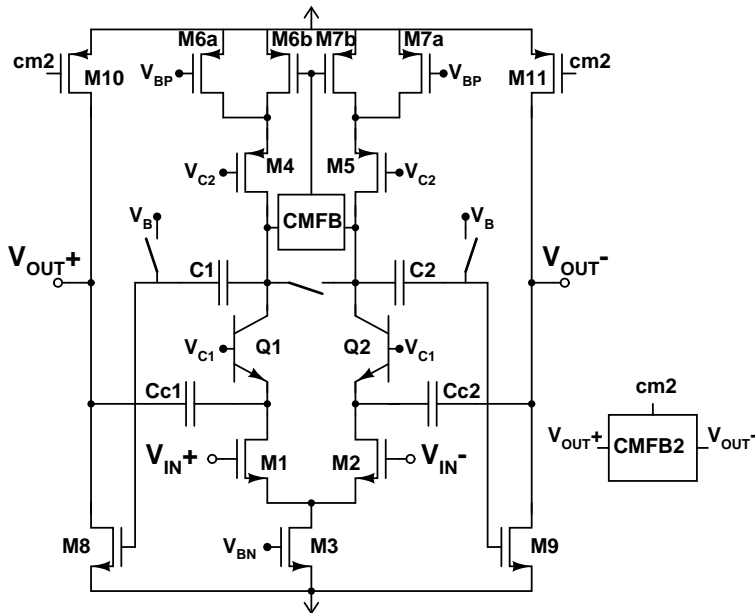
Figure 12.30 High-speed BiCMOS opamp.

smaller permissible signal amplitude. The total sampling capacitance (2C) is 10 pF, which leaves enough margin to the target resolution for the noise contribution of the opamp and the ADC.

12.5.3 Opamp

From the opamp, high gain, wide bandwidth, and low noise are required simultaneously. Furthermore, in order to obtain a high signal-to-noise ratio with low supply voltage, it is important to maximize the opamp output voltage swing, which makes the utilization of a rail-to-rail output stage almost a necessity. The need for a high DC gain excludes the possibility of resorting to the traditional Miller topology and thus the options are a three-stage architecture or a two-stage opamp with a high-gain first stage. The latter is chosen because of easier compensation and potentially higher speed.

The opamp architecture, shown in Figure 12.30 and already discussed in Section 7.3.2.2, is based on a telescopic input stage and a rail-to-rail output stage.

12.5.4 Switches

The input switches of the S/H circuit are realized with the double-side bootstrapped circuit introduced in Section 6.4.3.4. Thanks to the triple-well process, the switch transistor bulk can be made to track the signal during the tracking phase. This yields a remarkable improvement in linearity . Hold-mode feedthrough via the switches is suppressed by the canceling technique proposed in Section 6.4.3.5.

The sampling switches are also realized with bootstrapped circuits, which helps to minimize the size of the switch transistor as well as to reduce the signal-dependent charge injection due to the small voltage swing over the switch on-resistance.

The common mode settling of the S/H circuit (or a pipeline stage) need not be as fast as the differential settling, since the error caused by incomplete settling is attenuated by the common mode rejection ratio. However, if the common mode voltage is still changing at the end of the hold phase, signal-dependent on-resistance of the feedback switches may result in a differential error voltage. Thus, bootstrapping is also employed in the feedback switches, as well as the input switches and the feedback switches of the two first pipeline stages.

12.5.5 Clock Buffer and Clock Generator

In IF-sampling the signal-to-noise ratio easily becomes limited by jitter, because the SNR degradation resulting from jitter is proportional to the rate of signal change. For example, a 75-dB SNR at 200 MHz requires a jitter smaller than 141 fs.

Even if the external clock source was ideal (jitter-free), on-chip clock buffering can easily add more jitter than allowed. As is known from ring oscillators, the jitter is proportional to the total delay without depending on the number of delay elements used. Thus, in this design the target is to minimize the delay from the clock pin to the sampling switch by making the buffer chain as short and fast as possible.

The buffer utilized is shown in Figure 12.31. It has a differential input and internal common mode voltage biasing, which requires the clock signal to be externally DC decoupled. The buffer consists of two stages, a differential pair first stage and an inverter second stage. To avoid any additional delay the sampling clock (ϕ) is not generated with a non-overlapping clock generator, but taken directly from the buffer output. The signals needed to control the other switches and the opamp are made with a clock generator, though.

The first stage of the ADC samples the S/H circuit output with a signal whose edge falls before the sampling clock of the S/H circuit rises. Since the sampling clock cannot be delayed, the clock edge needed by the ADC is generated with a DLL which
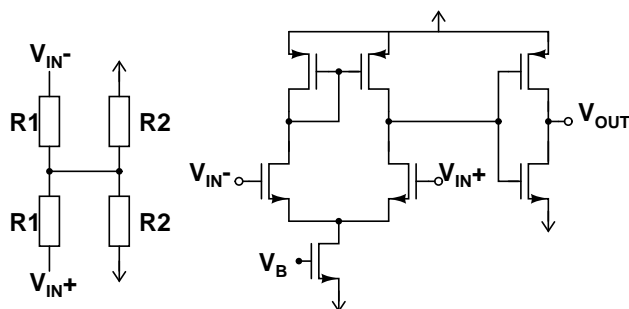
Figure 12.31 Clock buffer.

is locked to the sampling clock.

12.5.6   DLL

The architecture of the DLL is shown in Figure 12.32. The variable delay is adjusted to be a half of the clock cycle. As a result, the taps of the delay line provide evenly-spaced clock edges between the falling and the rising edge of the incoming clock signal.

The circuit consists of two voltage-controlled 16-element delay lines, the first of which belongs to the primary loop, which is the only active loop in the locked state. The second delay line is used to assure locking to the correct phase. It is fed with pulses instead of a continuous clock waveform, which allows the detection of the conditions where the primary loop is outside its phase capture range and tries to lock to on a wrong clock phase. The coarse control logic generates $up2$ and $down2$ pulses, which force the primary loop into the correct range. Once there, the coarse control becomes inactive.

The differential delay element is based on a current-starved two-inverter cell, which is controlled with two voltages. The loop filter is built around a differential gm-cell.

12.5.7   Self-Calibrated Pipeline A/D Converter

Self-calibration is applied to the first two 2.5-bit stages, while the 9+2-bit back-end pipeline ADC employs 1.5-bit stages. A small state machine, which generates the MDAC and S/H control signals during the calibration, and the RSD correction adders are also implemented on-chip. The 15 uncalibrated output bits, along with the raw outputs of the calibrated stages and three control signals, are fed out of the chip to the

Figure 12.32 Block diagram of the DLL.

external calibration logic and coding circuitry. The ten reference voltages needed in the coarse A/D conversions of the stages are generated with an on-chip resistor string, while the references for the capacitive MDACs are from off-chip voltage sources.

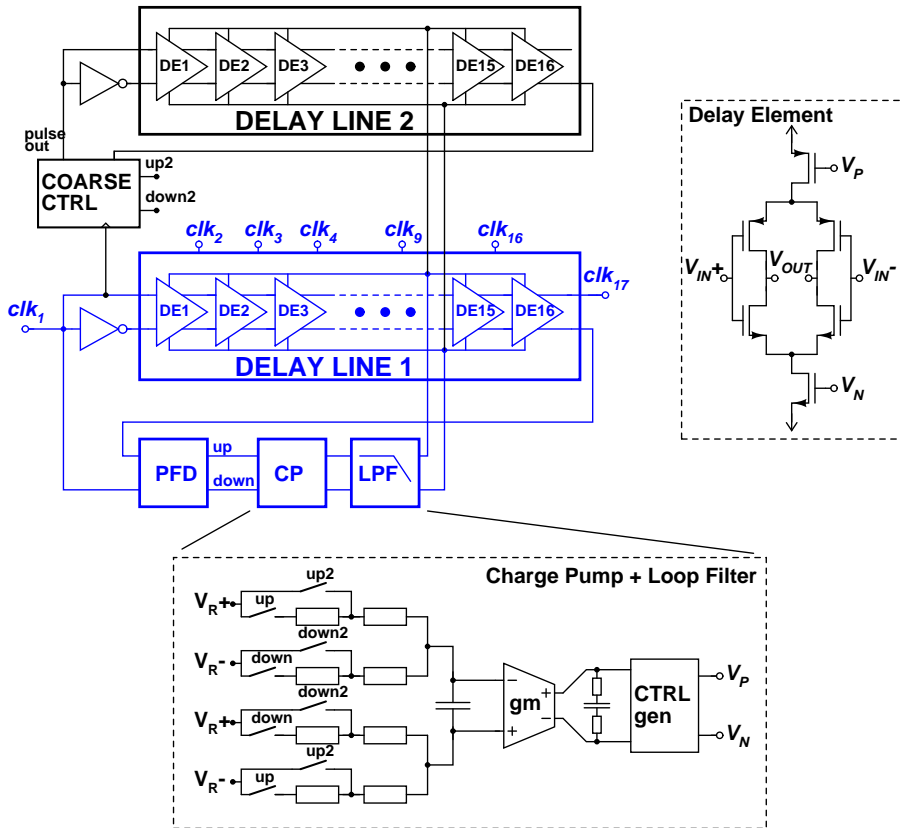Partitioning of the resolution is chosen on the basis of the fact that a high-resolution stage in front of the pipeline ADC provides linearity improvement as well as power savings in the subsequent stages (see Section 4.5.6). However, each additional bit in the first stage halves the opamp's feedback factor and doubles the number of comparators and their accuracy requirement. The same benefits are achieved by having two medium-resolution front-end stages, the amplifier and comparator specifications still being reasonable. The back-end pipeline should have the minimum stage resolution, so as to minimize power consumption. On the basis of simulations on a behavioral pipeline ADC model, the target 13-bit linearity can be achieved with two calibrated 2.5-bit stages in front of a 1.5-bits/stage back-end.

The specifications for the opamps used in the 2.5-bit and 1.5-bit MDACs are quite different. The 2.5-bit stages have a small feedback factor, which leads to a large open loop GBW requirement for the opamp. On the other hand, the opamps do not necessarily need to be stable in the unity gain feedback, since they are not auto-zeroed, which relaxes the phase margin specifications and thus makes it easier to fulfill the GBW requirements. Also, the DC gain has to be larger in the front-end stages, while the accuracy requirements scale down toward the LSB stages. The 1.5-bit stages have a larger feedback factor and a smaller DC gain requirement, but, simultaneously, higher stability specifications. The opamp topology used in the pipeline stages and shown in Figure 12.30 is similar to the one used in the front-end S/H circuit. The first two stages have similar opamps, which differ from the opamp used in the S/H circuit only in having a smaller compensation capacitor. The 1.5-bit stages all have identical opamps, where the bias current and device sizes are scaled down by a factor of four from the opamps of the first two stages. In the first pipeline stage 2.5-pF unit capacitors are used, resulting in a total sampling capacitance of 10 pF. In the second 2.5-bit stage and in the back-end pipeline stages the unit capacitors are scaled down to 1.0 pF.

The sub-ADCs of the 2.5-bit, 2-bit, and 1.5-bit pipeline stages are of the flash type and consist of six, three, or two comparators, respectively. The comparators drive the switches of the MDAC and a small decoding logic, which converts the thermometer code into binary output. In the calibrated stages a multiplexer is added to select between the normal MDAC switch control signals and the state machine-produced calibration signals. The low stage resolution, together with the RSD correction, allows the use of non-DC-power-consuming dynamic comparators. The sub-ADCs utilize the same differential pair dynamic comparator (Figure 12.20) as the parallel pipeline
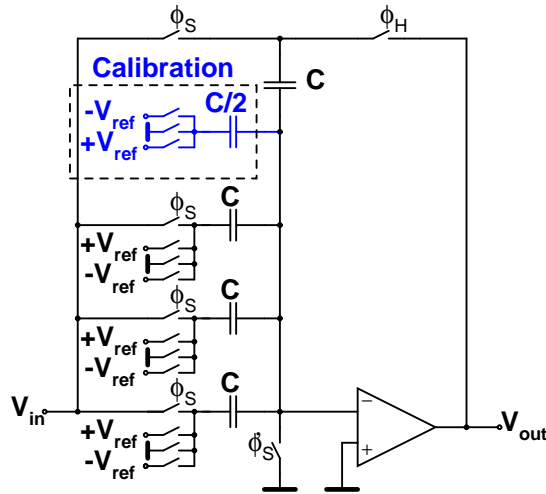
Figure 12.33 The MDAC employed in calibrated 2.5-bit stages.

ADC described in the previous section.

12.5.8   Calibration Circuitry

To meet the 13-bit resolution requirement with a high yield the effects of capacitor mismatch, low opamp DC gain, and reference voltage mismatch have to be compensated for. A widely-used method for 1-bit/stage pipeline ADCs is a digital self-calibration technique, where the discontinuities in the transfer function of each pipeline stage are serially measured using the back-end pipeline, giving correction coefficients to be added to the ADC output during normal operation. In the calibration method employed, developed from [82], the error attached to each reference unit capacitor is measured separately with the back-end stages and, following the switching scheme of the stage, correction coefficients for the stage output codes and the offset can be cumulatively calculated from these measurement results.

A simplified schematic of the 2.5-bit MDAC is shown in Figure 12.33. It differs from a conventional capacitive MDAC in that it has an extra calibration capacitor, the size of which is half of the unit capacitance. The stage under calibration samples the signal ground voltage, and in hold mode the capacitor being measured is connected to the positive or negative reference voltage, while the extra calibration capacitor is connected to the opposite reference voltage and the other capacitors to the ground. The resulting deviation of the output from the ideal $\pm V_{ref}/2$ is measured with the back-end pipeline. Similarly, the error of the calibration capacitor is measured and
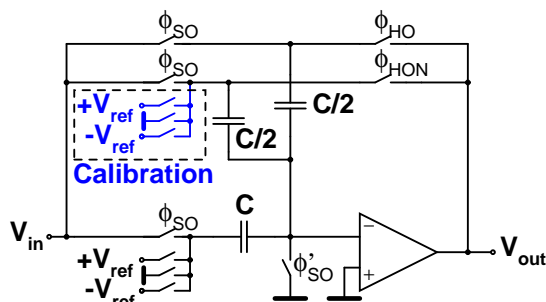
Figure 12.34 A 1.5-bit MDAC with the option of changing the gain from 2 to 4.

taken into account in the calculations.

Without this extra capacitor, the stage output during the measurements would be $\pm V_{ref}$ or 0, depending on whether one capacitor or a difference between two capacitors is measured. Measuring the differential capacitor errors (compared to the feedback capacitor) does not take account of the error resulting from the finite opamp gain, while performing the measurement in the vicinity of $\pm V_{ref}$ easily saturates the back-end stages in the presence of even small mismatches. The extra capacitor eliminates both of these problems by shifting the measurement of the capacitor values in the same voltage range—to the vicinity of $\pm V_{ref}/2$, where the transfer function steps are in normal operation.

As the output of the second stage during the calibration falls in a narrow voltage range near ± 0 or $\pm V_{ref}/2$, the first stages of the back-end pipeline are actually used as amplifiers, which makes possible the enhancement of the accuracy of the measurement by a factor of four, simply by doubling the inter-stage gain of the first two 1.5-bit stages. This is accomplished by halving the MDAC feedback capacitor during the calibration, as shown in Figure 12.34. Normally, the two halves are tied together, but during calibration one half is connected in parallel with the reference capacitor. Two extra stages are added to the back-end to get even more resolution when measuring the calibration coefficients, as well as to reduce the truncation error in the calibration adders during normal operation. The calibrated output is fixed to 13 bits.

The calibration algorithm is implemented in VHDL. The design allows the inclusion of the calibration logic on the same chip as the ADC or its realization it with an FPGA, which is combined with the ADC chip on the circuit board level. In this prototype the latter approach is used because of its flexibility. The calibration state machine controls the reference voltage switching in the MDACs of the first two stages during the calibration hold phases. In addition, turning off the S/H circuit and switching to the gain-of-four mode in the third and fourth stages are also controlled by the state
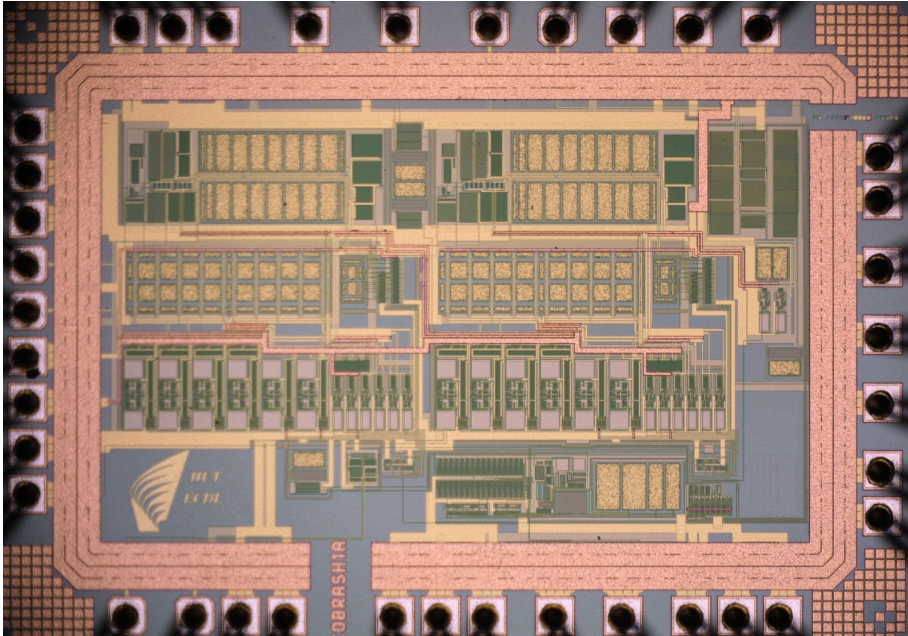
Figure 12.35 Die micrograph of the S/H prototype.

machine.

The input signals for the calculation and coding logic are the RSD-corrected digital words from the converter chip, the raw bits of the first and the second pipeline stage, the state-indicating bits, and an external reset signal. The seven calibration coefficients per stage are calculated as an average of four measurements and stored in a memory. During normal operation calibration coding is simply the addition of two coefficients, which are selected according to the raw bits of the two first stages, to each uncalibrated ADC output word.

12.5.9 Measurements of the Front-End

The front-end S/H circuit was first processed as a stand-alone prototype to allow its characterization without the ADC and the development of the measurement setup while the chip with the ADC was still in the fabrication process. The chip was fabricated on a 0.35-μm BiCMOS technology with silicon-germanium npn transistors and metal-insulator-metal capacitors. Figure 12.35 shows the die micrograph.
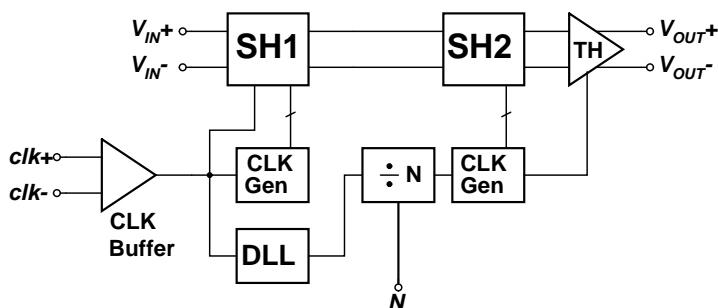
Figure 12.36 On-chip test setup.

12.5.9.1 On-Chip Circuitry

The on-chip circuitry for the beat frequency test is shown in Figure 12.36. The second S/H circuit is followed by a track-and-hold circuit, since the output of the S/H is valid for only one half of the clock cycle and is being reset to zero during the other half. The T/H circuit tracks the valid S/H output and holds it over the reset phase. The circuit is simply a switch and a capacitor followed by a linearized source follower buffer similar (except that the resistor is replaced with a current source) to the one employed in the S/H circuit shown in Figure 5.5.

The clock is brought in as a differential signal, which is first amplified and converted to a single-ended form. The resulting signal is used directly to control the sampling switches of the first S/H (the device under test) and as an input for the clock generator of the first circuit. The clock for the second circuit is generated from the incoming clock signal by dividing it with a programmable synchronous counter ($N \in [2, 17]$). The phase of the divided clock signal is locked properly to the sampling clock with the aid of the DLL circuit.

12.5.9.2 PCB

The test board is a 2-layer PCB in which the bottom layer is used as a ground plane. Four versions of the test board were made, one with a socket for the chip, and three where the prototypes were directly soldered to the board. These boards have different types of input circuitry, one being targeted on low-frequency measurements and the other two measurements with an IF input signal, the difference being in the clock input; one uses external clock source while the other has a crystal oscillator on the board.

The low-frequency input structure is shown in Figure 12.37. There, a fully differential discrete opamp is used for buffering the signal coming from the signal source.
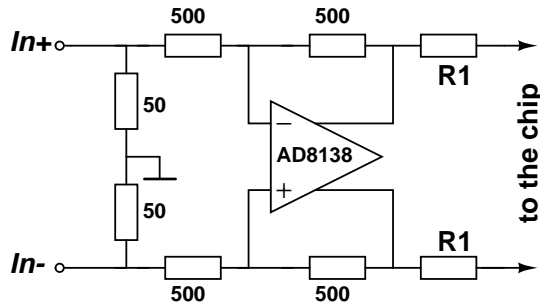
Figure 12.37 Input buffer for low frequency measurements.
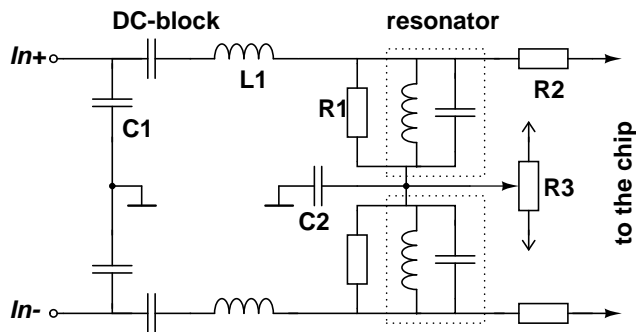
Figure 12.38 Input structure in IF measurements.

The purpose of resistors R1 is to attenuate the glitches produced by the switched-capacitor load. Initially, attempts were made to have the buffer circuit on a separate board, but the arrangement did not work well, probably as a result of too-high inter-connect parasitics.

At the 190-MHz IF frequency the opamp buffer solution is not applicable, because opamps with adequate bandwidth and linearity are not available. The current spikes resulting from loading the switched sampling capacitor have their energy at the clock frequency and its multiples. Thus, providing a low impedance path to ground at these frequencies facilitates the job of the signal source. This is realized by putting an LC resonator in parallel with the circuit input and tuning it to the IF frequency. The drawback of this solution is that it is an inherently narrow band, providing only a bandwidth of a few megahertz around the resonant frequency (the bandwidth, of course, depends on the Q value). Since the load current spikes have a large common mode component, the resonator cannot be inserted between the complementary input signals; instead,
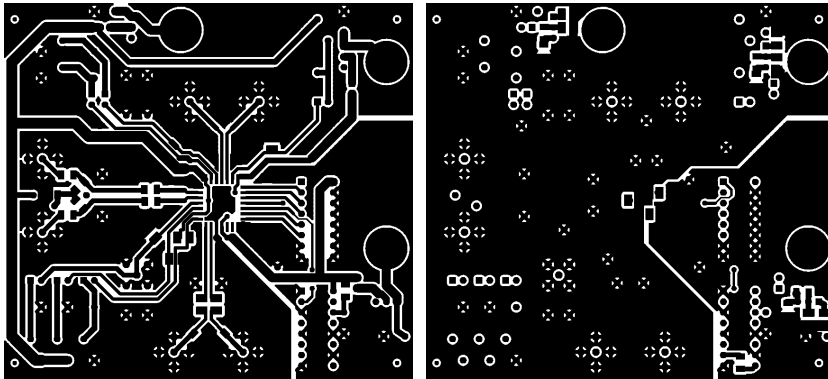
Figure 12.39 Two-layer PCB used in IF measurements.

both the inputs require a separate resonator against the signal ground.

The input circuit, shown in Figure 12.38, also includes matching elements (C1, L1, and R1), a DC-block, and common mode voltage adjustment (R3, which is bypassed with C2 at signal frequencies). The two-layer PCB used in IF measurements is shown in Figure 12.39.

12.5.9.3 Equipment

The input signal from the signal generator is first filtered with discrete LC filters in order to get rid of the harmonics. In the IF measurements a 190-MHz band-pass SAW filter was also tried. After filtering, the differential signals are generated with a power splitter.

The differential beat frequency output is converted to a single-ended form with a traditional instrumentation amplifier circuit constructed from three low-distortion opamps. The instrumentation amplifier has a 50-Ω driving capability. To achieve low distortion the gain of the amplifier is only 0.25 (to the 50-Ω load).

The clock input is driven directly from a sinusoidal signal source. For lower phase noise and jitter it is advantageous to take the clock from a crystal oscillator, which was also tested at the 50-MHz clock frequency.

12.5.9.4 Functionality

The S/H circuit and the on-chip measurement setup work in a functionally correct way with the nominal bias values. The DLL stays locked from ~ 25 MHz to ~ 120 MHz and the lock range can be somewhat shifted by changing the DLL bias current. The on-chip measurement setup works well up to a 60-MHz clock frequency, after which
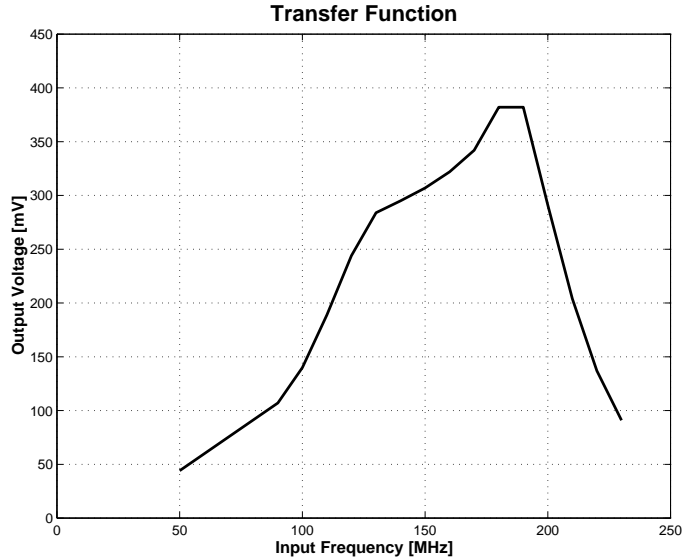
Transfer Function

Figure 12.40 Measured resonator transfer function.

the observed results suggest that there are some timing problems at the interface of the two cascaded S/H circuits. The exact frequency where the problems begin depends on the bias settings and the individual sample.

12.5.9.5   Problems and Difficulties

As described earlier, there were two different approaches to driving in the signal, opamp buffer at low frequencies and resonator at IF. The opamp buffer works well at low frequencies (10 MHz and below), but at 50 MHz it clearly limits signal purity. In between 10 and 50 MHz it is difficult to say whether the performance is limited by the chip or the buffer.

In the IF, the resonator ideally (according to simulations) almost completely isolates the current spikes from the signal source. In reality, however, it seemed to be very difficult to construct a resonator with a high quality factor, which was probably due to PCB parasitics not included in the simulations. The low Q resonator attenuates the current spikes, but probably not enough. Thus, the results obtained at 190-MHz IF are likely to be partially limited by the purity of the input signal rather than the circuit itself. This assumption is supported by the fact that the bias adjustments did not have a strong effect on the results.

The measured transfer function is shown in Figure 12.40. The resonance peak is not as sharp as predicted by simulations and neither was its location correct with the

nominal component values (L=3.1 nH, C=220 pF). By reducing the resonator capacitance to 161 pF the resonant frequency was shifted to the desired 190 MHz.

In the IF measurements the level of the second harmonic was found to be rather high and any of the chip's voltage or current bias settings has virtually no effect on it. The only thing that seemed to have an effect, which indeed was considerable, was the amplitude of the clock signal. The most probable explanation is that there is some coupling from the input signal to the clock. As shown in Section 8.2 the coupling results in a spurious signal at twice the signal frequency. By changing the amplitude of the sinusoidal clock the slope changes and the observed effects seem to be in line with the theory. The effect of signal frequency could not be tested because of the narrow bandwidth of the input circuitry.

It was never found out where the coupling actually happened. Adding decoupling capacitors in various places in the PCB and improving the return current paths around the signal and the clock traces in the PCB did not have a significant effect. So the coupling probably does not happen on the PCB. It may occur between the bond wires or on the chip. One possible on-chip coupling mechanism is through modulation of the supply voltage.

12.5.9.6   Low Frequency Results

At low frequencies the circuit showed such low distortion that when the signal amplitude was ~2.5 dB below the full scale or less the harmonics disappeared below the spectrum analyzer noise floor, which was then at the 75–78 dBc level.

12.5.9.7   IF Results

The levels of the second and third harmonics are plotted as functions of the output signal amplitude at the 50 and 60-MHz clock frequencies in Figures 12.41–12.44. As mentioned earlier, the level of the second harmonic depends almost solely on the amplitude of the clock signal. From the 50-MHz results it can be seen that, when taking account the 6 dB difference between the two gain modes, the second harmonic for a given input amplitude is the same regardless of the S/H circuit gain. The level of the harmonic decreases 6 dB per decade faster than the input amplitude, which supports the theory of clock contamination.

It was clearly seen that the phase noise with the crystal clock was significantly lower than with a signal generator clock (reference locked to the input signal generator). The level of the second harmonic, on which only the clock amplitude had some effect, was higher than in the measurements with the signal generator as the clock.
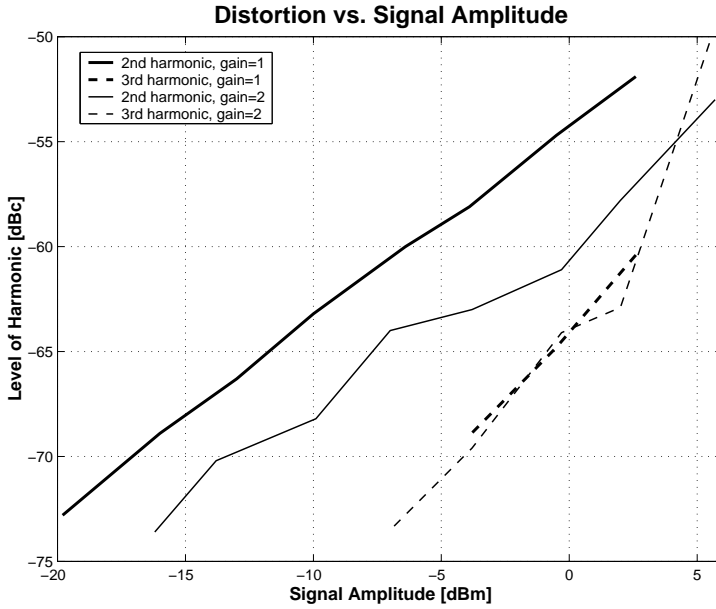
Figure 12.41 Distortion at the 50-MHz clock frequency. The input frequency is 190.2 MHz and the full-scale amplitude corresponds to +3.8 dBm.
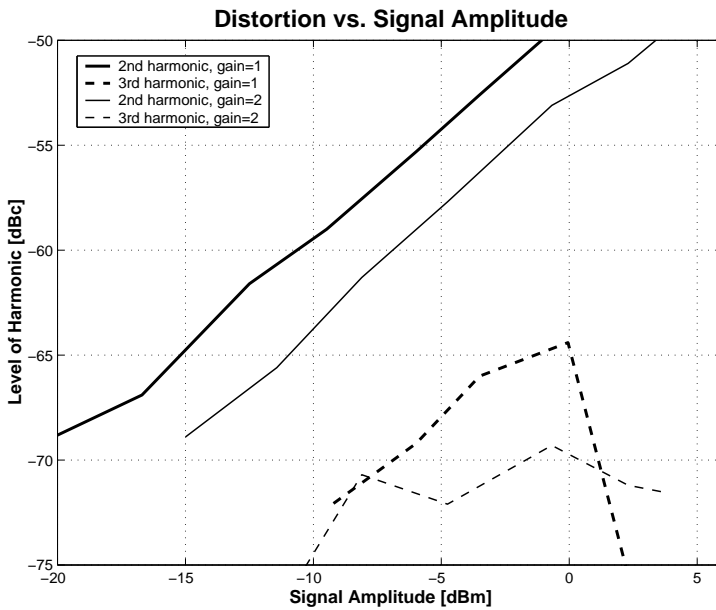
Figure 12.42 Distortion at the 50-MHz clock frequency taken from an on-board crystal oscillator. The input frequency is 190.1 MHz and the full-scale amplitude corresponds to +3.6 dBm.
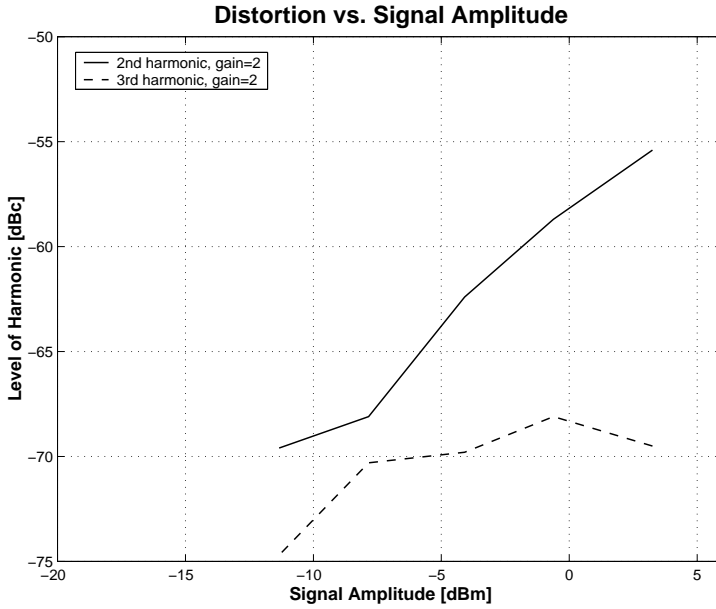
Figure 12.43 Distortion at the 50-MHz clock frequency taken from an on-board crystal oscillator. The input frequency is 140.1 MHz and the full-scale amplitude corresponds to +3.6 dBm.
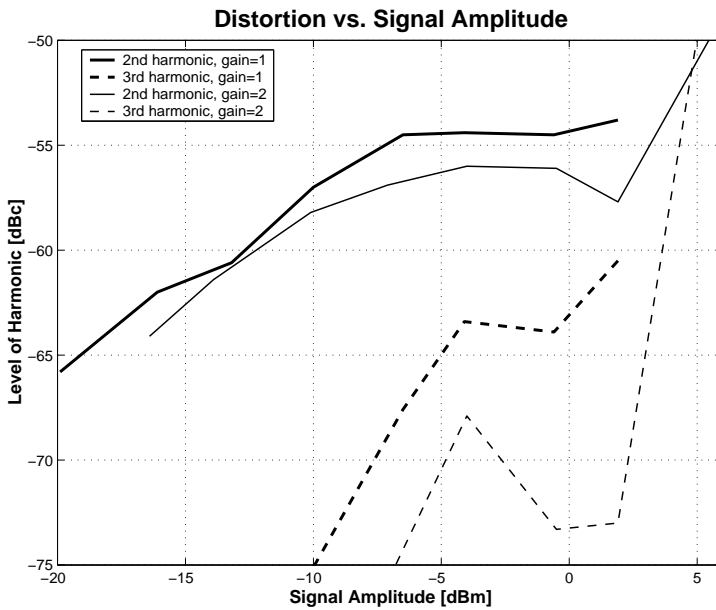
Figure 12.44 Distortion at the 60-MHz clock frequency. The input frequency is 190.2 MHz and the full-scale amplitude corresponds to +3.8 dBm.
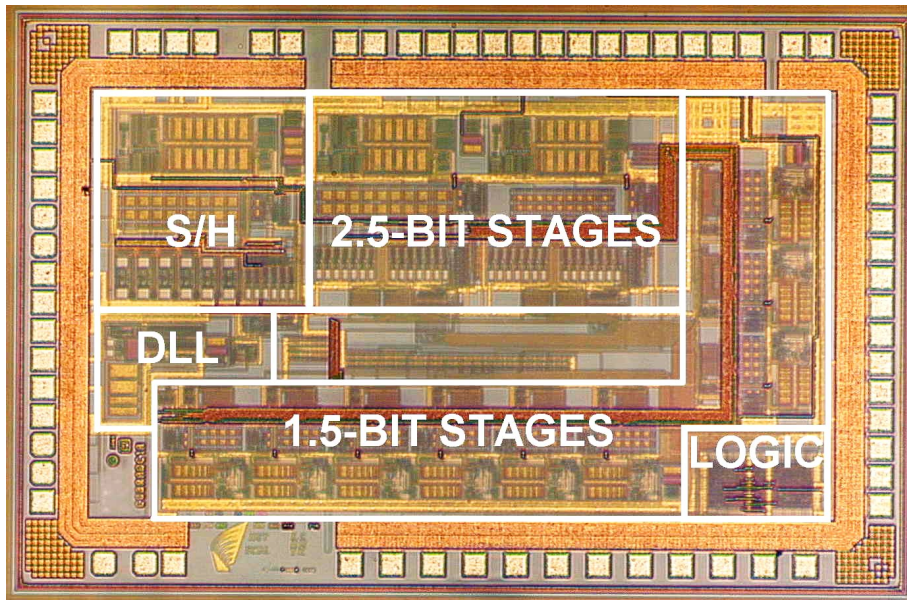
Figure 12.45 13-bit 50-MS/s pipeline ADC die micrograph.

This is because the clock amplitude could not be adjusted to minimize the distortion and probably because the clock signal was single-ended unlike the one obtained with the generator.

Some form of filtering is always needed to remove signal source harmonics. All the IF results presented were obtained with a 200-MHz lowpass LC filter. A bandpass SAW filter (single-ended) and combinations of the SAW and the LC were also tested, but the results were practically unchanged. The problem with the SAW filter is its high attenuation ($\sim$9 dB), which prevented measurements with signal amplitudes close to the full scale, because the signal generator amplitude was already almost at its maximum without the SAW device.

12.5.10 ADC Experimental Results

The circuit containing the whole system shown in Figure 12.28 was fabricated with the same 0.35-μm BiCMOS technology as the stand-alone front-end, and the chip was packaged in a 52-pin VFQFPN package. The benefits of this almost chip-scale package are small parasitics and also a low ground inductance and good thermal conductivity, which are thanks to the die attachment to a large pad exposed through the package.

The prototype was measured using a 4-layer PCB, the FPGA being on a separate

Table 12.3 Summarized ADC performance.

Resolution	13 bits
Sample Rate	50 MS/s
Input Range (differential)	3.8 V
Input Bandwidth	> 200 MHz
DNL / INL (calibrated)	± 1.0 / ± 3.0 LSB
SFDR (@ 200 MHz)	76.5 dB
Supply Voltage	2.9 V
Power Dissipation	715 mW
Die Area	6.0 mm^2
Technology	0.35-μm BiCMOS (SiGe)

board. A common ground plane was used for the analog and the digital circuitry. The clock signal from an on-board 50-MHz crystal oscillator was connected to one of the differential clock inputs, while the complementary input was grounded near the oscillator. The arrangement for bringing in the IF signal was identical to the one used with the stand-alone S/H prototype.

The first measurements revealed that there was a timing error in the digital delay line used for aligning the output bits of the pipeline stages. As a result the output contained a large number of seemingly random bit errors. It was found that these errors could be minimized, but not totally eliminated, by lowering the supply voltage from the designed 3.0 V to 2.9 V and performing the testing at a temperature of $+5°C$. Due to the remaining bit errors the noise floor was high, limiting the SNDR to 55–60 dB. On the other hand, the linearity did not seem to suffer significantly.

The static linearity was measured with a 195.2-MHz signal and calculated using the code density test. Figure 12.46 and shows the results before and after the calibration cycle. The calibration improves the maximum INL from ± 7.1 LSB to ± 3.0 LSB, the DNL being within ± 1.0 LSB in both cases.

The SFDR measured with a -1-dBFS (3.4 Vpp differential) signal was better than 73 dB in all frequencies in the range from 190 MHz to 200 MHz. An example spectrum with a 76.5-dB SFDR is shown in Figure 12.48. The effect of calibration can be seen by comparing it with Figure 12.47, where the same measurement is repeated without the calibration. The measured power consumption from the 2.9-V supply was 715 mW. Table 12.3 summarizes the ADC performance.

The measurements showed a greatly improved second harmonic at IF compared to the stand-alone version of the S/H circuit. The possible reasons for the reduced clock contamination are the larger separation of the input and the clock pads, smaller bond wire and package inductances thanks to the VFQFPN package, and reduced coupling
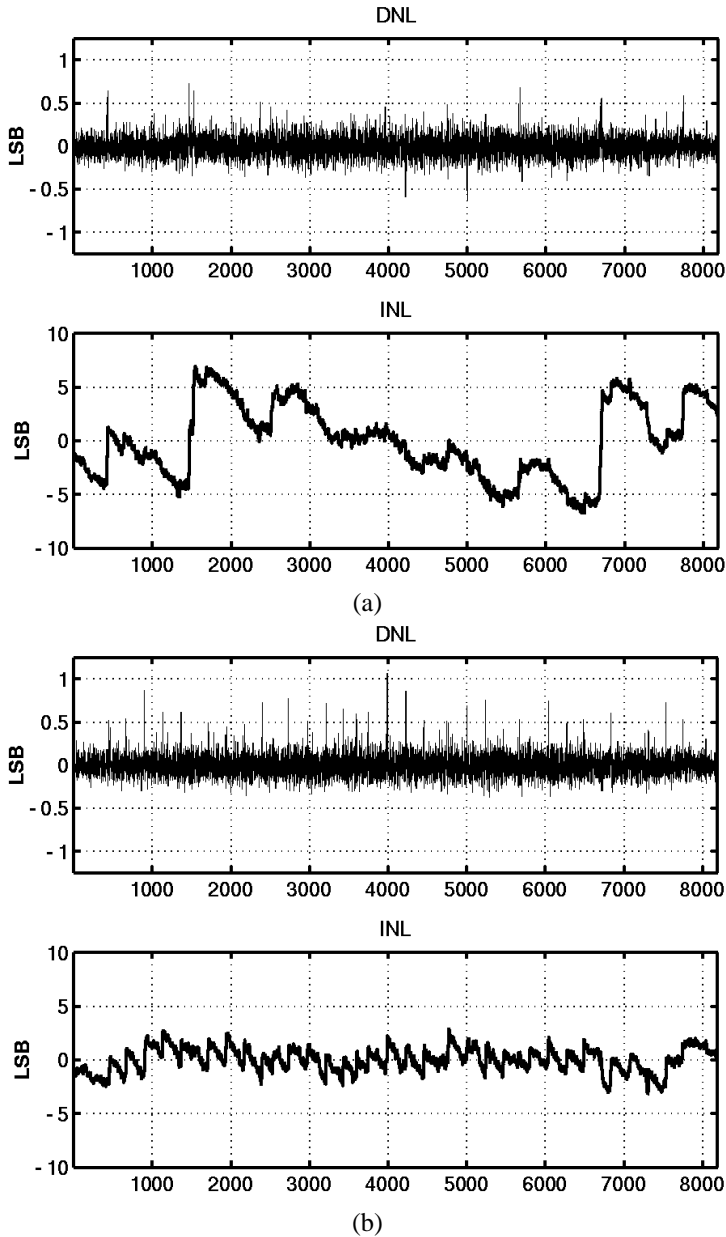
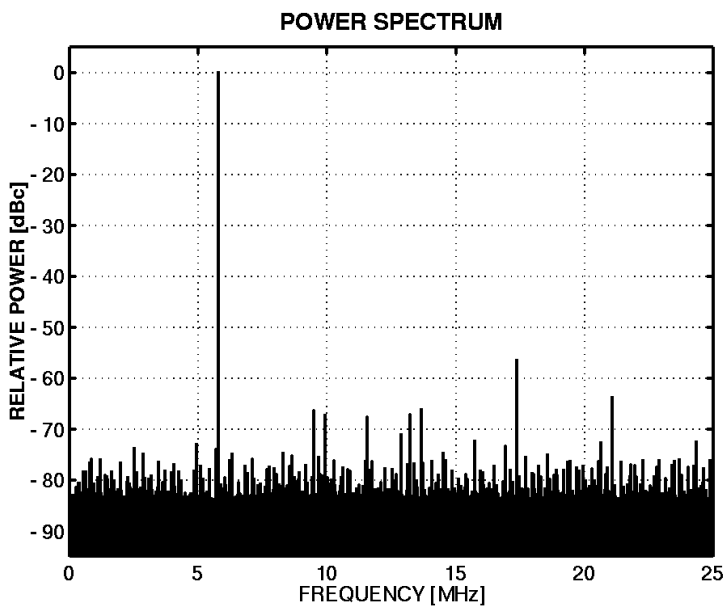**Figure 12.46** DNL and INL (a) before and (b) after calibration.

Figure 12.47 A spectrum measured with a 194.2-MHz, −1-dBFS signal before calibration.
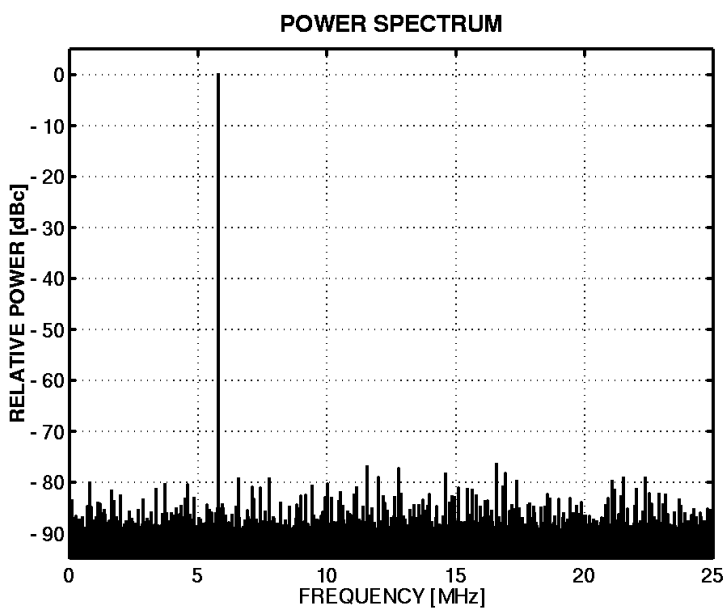
Figure 12.48 A spectrum measured with a 194.2-MHz, −1-dBFS signal after calibration shows a 76.5-dB SFDR.

on the board level resulting from 4-layer PCB.

12.6 Deglitcher for Current Steering DACs

12.6.1 Introduction to Current Steering DACs

Traditionally the applications of high-speed DACs have been in video and computer graphics applications, but recently the migration to wideband wired and wireless telecommunication standards and the evolution of radio transmitter architectures toward the software-defined radio have created a need for high-speed, high-resolution telecommunication DACs.

In the past the research and development of DACs have been heavily concentrated on improving the static (DNL, INL) and, to some extent, the time domain specifications (settling time, glitch area), almost totally neglecting spectral purity and other frequency domain characteristics that are essential in telecommunication devices.

Practically all high-speed DACs are based on the current steering architecture, one of the main reasons for this popularity being its capability of driving resistive loads without buffering. A 14-bit static linearity has been achieved by using trimming, self-calibration [230], and even intrinsically [231]. A typical problem in these DACs is the rapid increase in harmonic distortion when the signal frequency is increased. This is mainly due to the glitches occurring at the code changes. The glitches are results of incoherent timing of the current switches, non-optimal shape of the switch control waveforms, and coupling of the digital signals to the analog output.

Attempts to reduce glitches include the use of latches to synchronize the switch controls, circuits to generate optimal control waveforms for the switches, and the use of return-to-zero-type output to suppress the output during the code changes [232]. The return-to-zero technique utilized in [233] yields a clear improvement in high-frequency SFDR compared to earlier reported DACs, but still has some limitations, such as the difficulty of providing large amplitudes to a low-resistance load, the complicated circuitry needed to handle signal dependent parasitics, and sensitivity to clock jitter, which is not relaxed, unlike in conventional DACs, when signal frequency is decreased. To alleviate the first two of these problems the same authors have proposed the track/attenuate technique [230], which is basically a switch put in parallel with the load to short the output during the DAC switching.

To avoid the jitter problem and signal attenuation it is possible to use a track-and-hold circuit as a deglitcher; the DAC is cascaded with a T/H circuit which tracks the DAC output when it is in steady state and holds a sampled voltage during DAC

Current Bias   5 Digital Trimming

Biasing

$I_{B,LSB}$ **1 : 256** $I_{B,MSB}$

LSB	**MSB**
8 Binary Weighted	**2 Binary Weighted**
Current Sources	**+**
	15 Nonweighted
	Current Sources

Digital
Input
14   **4 to 15**   15   1 • • • 8   1 • • • 17
      encoder

      10   **Latches**   25   **Current Switches**

              DAC Clk I_{DAC}+ I_{DAC}-

Clk+   **Clock**
Clk-   **Buffer**   **DG Clocks**   **Deglitcher**

                        I_{OUT}+ I_{OUT}-
                        Current Output
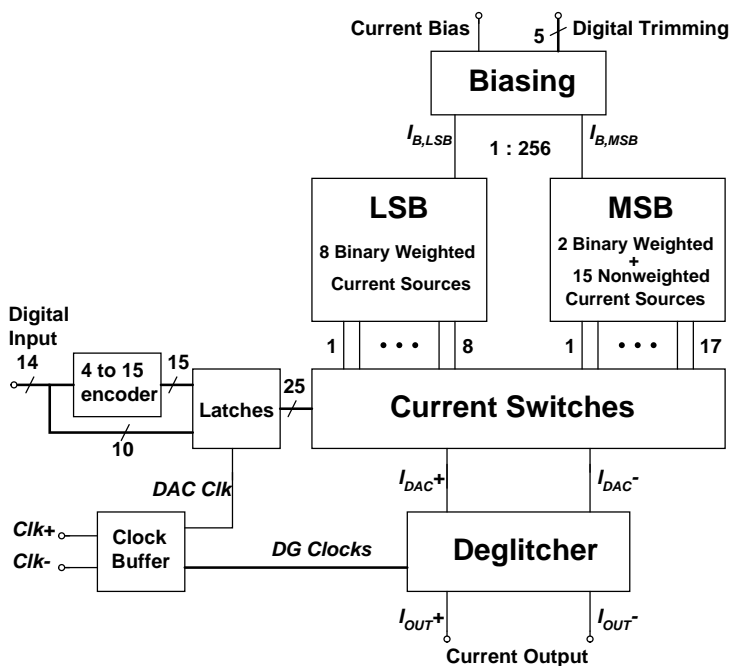
Figure 12.49 Block diagram of the DAC.

settling. Although the deglitcher does a good job of removing code-dependent glitches it typically cannot achieve as high a speed as a current-steering DAC and, furthermore, the voltage output provided by the T/H needs to be buffered in order to drive resistive loads.

The DAC presented here and published in [15] and [16] employs a deglitcher which is based on current mode circuitry. The output is provided in the form of a current which eliminates the need for a buffer. A high speed is achieved by employing parallelism. The circuit does not rely on matching, which makes it very robust and eliminates the need for calibration.

12.6.2   Circuit Description

12.6.2.1   Architecture

The core DAC, shown in Figure 12.49, is based on segmented architecture. The current sources are divided into two unit current source arrays, with 8 LSBs in one and 6 MSBs in the other. In the LSB array the current sources are binary-weighted and constructed of parallel unit sources distributed around the matrix to compensate for linear and center-symmetric process variations. In the MSB matrix the two least signifi-

cant bits are binary-weighted and the remaining four are formed with 15 unweighted sources constructed of four unit sources laid out in common centroid geometry. To reduce cumulative mismatch errors, the consecutive unweighted sources are selected in such a manner that if one source is constructed of transistors on the periphery of the matrix, the next one will have its transistors closer to the center, and vice versa.

The current bias for the LSB and MSB arrays is generated in a bias array from a single external reference current. The bias array is a current mirror that generates two currents, the LSB bias being 1/64 and the MSB bias 4 times the reference. The current ratio can be manually trimmed with an external 5-bit control signal.

The current switches are controlled with a 10-bit binary code and a 16-level thermometer code, which are synchronized with a latch stage before the switches.

The DAC output is connected to the deglitcher, which requires a 1.75-V voltage headroom; thus, the DAC has to fit within 1.25 V when a 3.0-V supply voltage is used. This has an effect on the sizing of the current sources and the DAC switches; the current sources are biased to a 1.2-V gate-source voltage to minimize the effect of threshold voltage variation in the available voltage headroom. The single-ended full-scale output current of the DAC is 10 mA.

The clock signal is brought into the chip in differential form to improve the noise rejection on the board and package level.

12.6.2.2 Deglitcher

The principle of the deglitcher is shown in Figure 12.50. It consists of four single-ended current mode sample-and-hold circuits and four switch pairs that operate in time-interleaved fashion. During one clock cycle the DAC differential output current is sampled into two current memories and the other two supply to the output the current which has been sampled in the previous clock cycle. This way the DAC is never directly connected to the output and the sampling phase is extended to cover nearly the whole clock cycle, maximizing the speed.

In the sampling phase the current memory is enclosed in a feedback loop that forces its current to be equal to the DAC output current. At the end of the phase the feedback loop is opened and the current gets sampled in the memory. In the hold phase the current memory acts as a current source, whose value is the one stored in the memory. In this phase the circuit is connected to the external load. Since the current value is set by feedback and kept unchanged when the circuit is connected to the load, no accurate matching is required between the current memories.

The clock waveforms for the circuit are shown in Figure 12.51. The DAC is

Digital input

Current Steering DAC

DAC clk

I_{OUT}^{-}

I_{OUT}^{+}

I_{DAC}^{-}

I_{DAC}^{+}

Clk+   Clk-   Clk-   Clk+   Clk+   Clk-   Clk-   Clk+

I1   I2   feedback   feedback   I3   I4

Clk2   Clk1   Clk1   Clk2

Current memory 1

Current memory 2

Current memory 3

Current memory 4
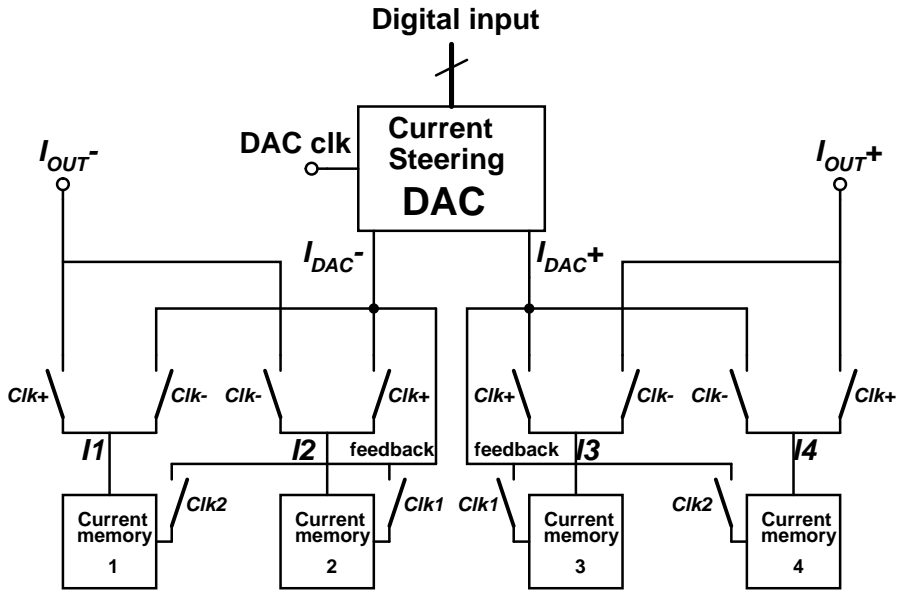
Figure 12.50 The deglitcher and clock signals.
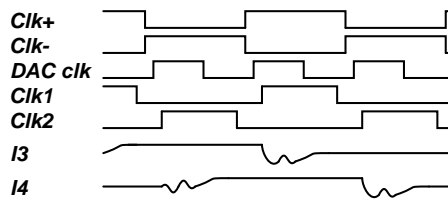
Clk+
Clk-
DAC clk
Clk1
Clk2
I3
I4

Figure 12.51 Clock signals.
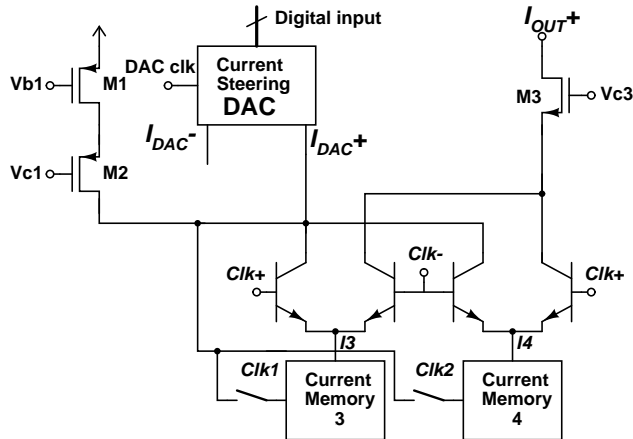
Figure 12.52 BiCMOS implementation of the deglitcher.

clocked at the full rate, while the deglitcher uses a set of half-rate clocks. The current switches are controlled with the complementary signals *Clk+* and *Clk-*, which have a 50% duty cycle. Signals *Clk1* and *Clk2* are non-overlapping clocks for the current memories. Two currents *I3* and *I4* are also shown to clarify the operation.

A more detailed implementation of the deglitcher is shown in Figure 12.52. There, the current switches are implemented with bipolar transistors and a cascode transistor is inserted between the current switch and the circuit output. In addition, a cascode current source is added in parallel with the DAC to bias the current memory. The base currents of the switch transistors—as long as the transistors in the switch pair are matched—are not a problem, since the switches are enclosed in the feedback loop when the current is sampled.

12.6.2.3 Current Switches

The bipolar current switches are driven with the circuit shown in Figure 12.53. It uses two voltages V_H and V_L, to which the bipolar transistor base is connected in the on- and off-phases, respectively. The switches S1 through S4 are bootstrapped MOS switches similar to the one in Figure 6.18. These switches are controlled with overlapping signals *d1* and *d2*, which, together with their complements, are shown in Figure 12.53. Also shown are the switch transistor base voltages, which cross near the high voltage level to avoid turning off both the transistors simultaneously, which would disturb their common emitter node and produce a glitch in the output. The crossing point depends on the overlap time and is adjusted to be too high rather than
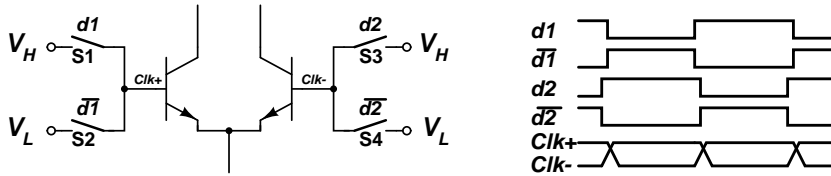
Figure 12.53 Current switch and its timing.
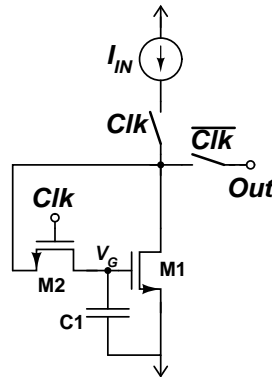
Figure 12.54 Basic current memory [235].

too low so as to guarantee desired operation under all process conditions.

12.6.2.4 Current Memory

Typically, the current memories are based on a MOS transistor whose gate capacitance is used as an internal storage element, while the memory input and output are both the drain current [234]. Such a basic circuit is shown in Figure 12.54 [235]. There, the transistor M1 acts as a voltage-controlled current source, the current of which is set equal to I_{IN} during the *Clk* high phase. This is accomplished by connecting the transistor in a diode configuration through the MOS transistor switch M2. As a result M1's gate voltage V_G settles to a value which makes the drain current equal to I_{IN}. When *Clk* goes low, M2 turns off and the voltage V_G is sampled in the capacitor C1, which can be the gate capacitance of M1 or a combination of the gate capacitance and an additional capacitor. Now M1 acts as a current source equal in value to I_{IN} at the sampling instant.

As regards high resolution applications, the most severe limitation of this circuit is the harmonic distortion originating from the sampling switch charge injection [236]. The nonlinear relationship between the gate voltage and the drain current results in a situation where even a constant charge injection from M2 produces harmonic distor-
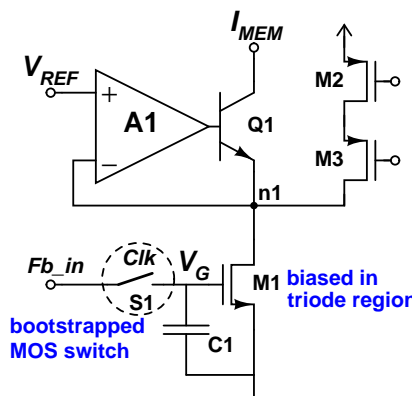
Figure 12.55 BiCMOS current memory.

tion in the output current. Moreover, the switch M2 operates against the voltage V_G, which makes the charge injection signal-dependent. Another problem is the limited output impedance of M1, which also distorts the output current if the output voltage during the hold phase does not match the voltage in the sampling phase.

The output impedance can be substantially improved simply by cascoding M1 with another device. Ways of reducing the effect of the charge injection range from using differential circuitry or dummy switches to the very accurate S^2I technique [237], where the sampling is done with two memory elements in two phases; first, a coarse sample is taken in the larger memory and in the next phase a correcting fine sample is taken in the other one. As a result, the remaining error is proportional to the size of the small fine memory. The disadvantage of this technique is the increase in sampling time introduced by the added second sampling phase.

In this design the approach taken to achieve 14-bit resolution is two-fold. First, the linearity of the current memory is maximized to make the circuit less sensitive to the constant charge injection, and second, the signal dependency of the charge injection is significantly reduced. Furthermore, when the current memory is linear, even an error linearly dependent on the signal can be tolerated, since it only affects the signal amplitude.

In [238] the signal-dependent charge injection is avoided by adding an opamp in the feedback loop to create a virtual ground at the drain of M1. The sampling switch is moved from M1's gate to the virtual ground, which makes its charge injection independent of the signal. The extra element in the feedback loop, however, inevitably increases the settling time. For this reason the approach taken in this design is to improve the switch itself.

The current memory is shown in Figure 12.55. The switch transistor gate-source voltage is made virtually constant by using bootstrapping. Neglecting the bulk effect, this makes the channel charge, as well as the injection resulting from its release, constant. The actual switch realization is based on the circuit presented in [136] and shown in Figure 6.18. The maximum gate overdrive yields a low on-resistance with a small switch transistor, minimizing the nonlinear parasitic capacitances, which are effectively in parallel with the memory capacitor.

Besides the channel charge, there is also charge redistribution in the gate overlap capacitance C_{Gol}. Since, when entering the off-phase, the switch gate is pulled to a constant voltage, the resultant signal-dependent error charge is $C_{Gol} \cdot V_G$. When the capacitances are constant (which is mostly true) and the current memory linear the error results in only a small change in signal amplitude. If necessary, this error could be avoided by switching the gate to a voltage that is the V_G properly buffered and level-shifted.

The high linearity of the current memory is based on the fact that the transistor M1 is biased in the triode, not the saturation, region. There, the drain current is given by

$$I_D = \frac{\mu C_{ox} W}{L} \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right].$$ (12.1)

Now, if the drain-source voltage is kept constant, the circuit is perfectly linear. To make the voltage on the drain as constant as possible M1 is cascoded with the bipolar transistor Q1, which has an inherently large gm, which is further boosted by using regulation. Regulating a bipolar cascode transistor, in contrast to a MOS transistor, does not improve the output impedance [239]. But, as already said, that is not the main reason for the regulation here. To achieve a high linearity it is necessary to bias the transistor M1 deep in the linear region. The cascode current source, consisting of M2 and M3, is for biasing M1.

The sampling speed of the deglitcher is determined by the time constants associated with the feedback loop. When the loop is broken at the gate of M1 (leaving the capacitor C1 on the output side) there is only one high impedance node (the DAC output), which justifies the use of the single pole approximation, giving the following gain-bandwidth product:

$$GBW = \frac{g_{m1}}{C_1 + C_{DAC}},$$ (12.2)

where g_{m1} is the transconductance of M1 and C_{DAC} the DAC output capacitance. Since the transconductance (together with the full-scale output current) determines the volt-
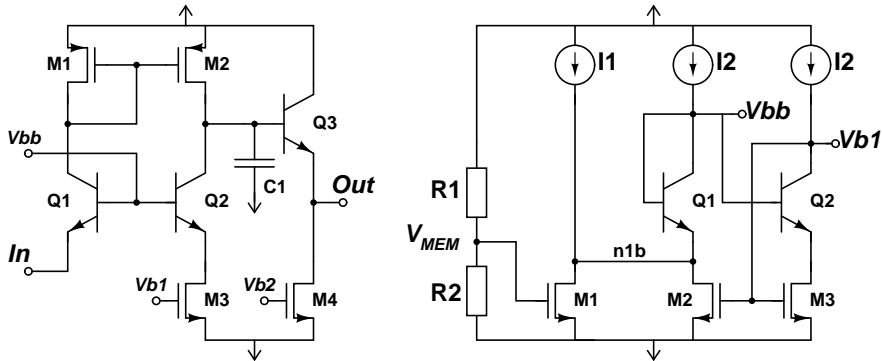
Figure 12.56 Regulation amplifier (left) and the bias circuit (right).

age swing on the memory capacitor and the capacitor value the sensitivity to charge injection and noise, there is a tradeoff between speed and accuracy. The sampling switch on-resistance and the other nodes in the loop produce non-dominant poles that affect the phase margin. These poles are given by

$$p2 = \frac{1}{R_{ON}} \cdot \frac{C_1 C_{DAC}}{C_1 + C_{DAC}}, \tag{12.3}$$

$$p3 = \frac{g_{m2}}{C_{n1}}, \tag{12.4}$$

and

$$p4 = \frac{g_{m,sw}}{C_{OUT,mem}}. \tag{12.5}$$

The auxiliary amplifier used in the regulated cascode transistor is shown in Figure 12.56. It consists of a transresistance input stage and an emitter follower buffer, which is needed to supply the large base current of the cascode BJT.

The current memory is biased via the regulation amplifier. The bias circuit is shown in Figure 12.56. It uses a scaled-down replica of the current memory (M1 and I1), the input voltage (V_{MEM}) of which is set at the desired level (1.5 V) with R1 and R2. This arrangement sets the drain voltage and, more or less, the transconductance of M1 correctly regardless of the process parameters and temperature. The drain voltage is used to generate the correct gate bias V_{bb} for the regulation amplifier.

12.6.2.5 Limitations

A potential problem in all circuits using time interleaved parallelism is mismatch between the parallel circuits. The current memory-based deglitcher, however, does not
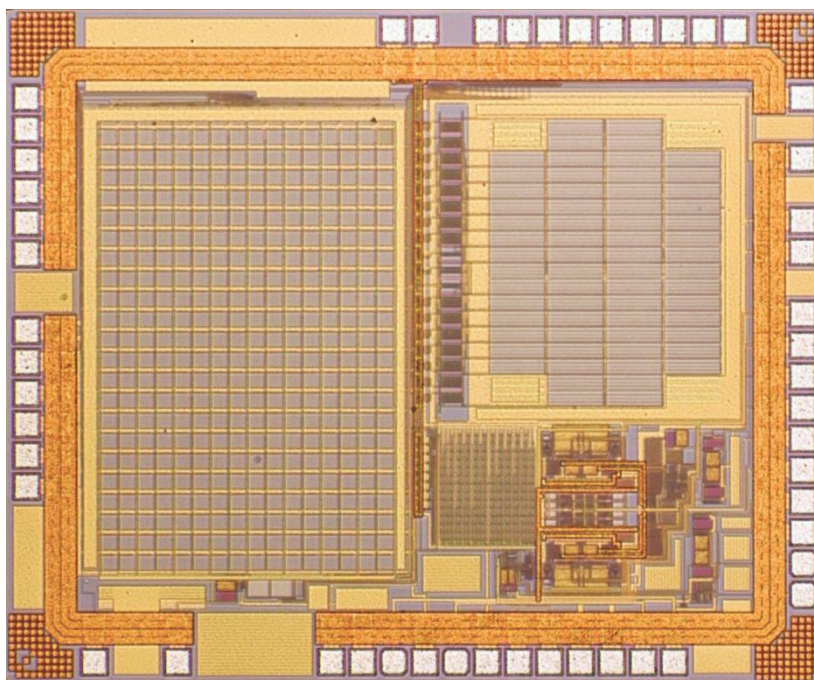
Figure 12.57 Photograph of the prototype 14-b DAC and the deglitcher.

rely on matching, thanks to the use of the current copying principle. The only parallel element not enclosed in the feedback loop in the sampling phase is the current switch. However, being constructed of bipolar devices, its matching should be adequate. Although the component matching is not a problem, the timing of the current switch can be. Any deviation from the 50% duty cycle will produce a spectral image of the signal around the half clock frequency. The error is signal frequency-dependent, getting worse as the frequency is increased. In most applications some amount of over-sampling is used, which moves the image outside the signal band and makes it tolerable up to some limit.

To minimize the error the half-rate clocks for the current switch control circuit are generated by dividing the full-rate clock with a carefully matched synchronous divide-by-two circuit constructed with a fully differential D-flipflop. For prototyping purposes a manual rise time control circuit was included in the signal path for fine-tuning the duty cycle if it turned out to be necessary.
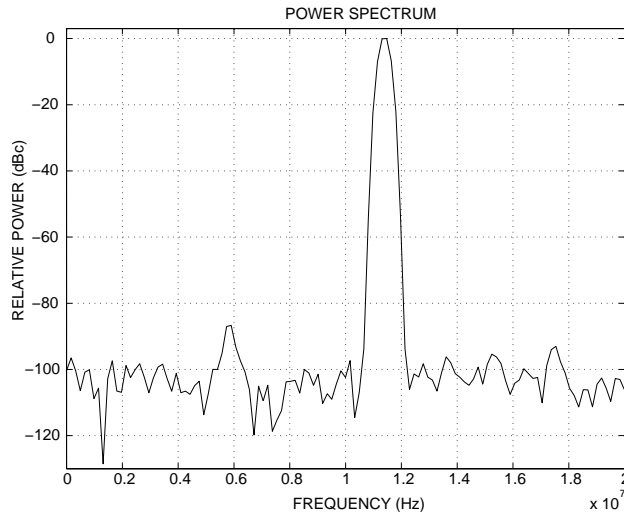
Figure 12.58 Simulated spectrum.

12.6.3 Simulations and Experimental Results

The circuit was designed using a 0.35-μm BiCMOS (SiGe) technology. The chip, a photograph of which is shown in Figure 12.57, occupies a total of 5.7 mm^2 of silicon area. The majority of the area is consumed by the DAC current sources, the deglitcher being only a small block on the lower right corner of the chip. A minor layout error in the first processed version, unfortunately, prevented any decent performance figures from being obtaining.

A simulated spectrum with an 11.4-MHz input signal and a 40-MHz clock frequency is shown in Figure 12.58. The highest spurious, the third harmonic, lies 86 dB below the signal and the level of the fifth harmonic is -92 dBc. No even order harmonics can be seen, thanks to the differential circuitry. The power consumption from a 3.0-V supply is 370 mW and dominated by the deglitcher with its 70% share.

12.6.4 Conclusions

The idea of using a current mode track-and-hold circuit as a deglitcher after a current-steering DAC has been proposed and demonstrated with a prototype circuit. The time-interleaved deglitcher uses the current copying principle, which makes it insensitive to component mismatch. The current memory developed achieves exceptionally high linearity thanks to a bootstrapped sampling switch and a transconductor constructed of a triode region MOSFET cascoded with a regulated bipolar transistor. Accord-
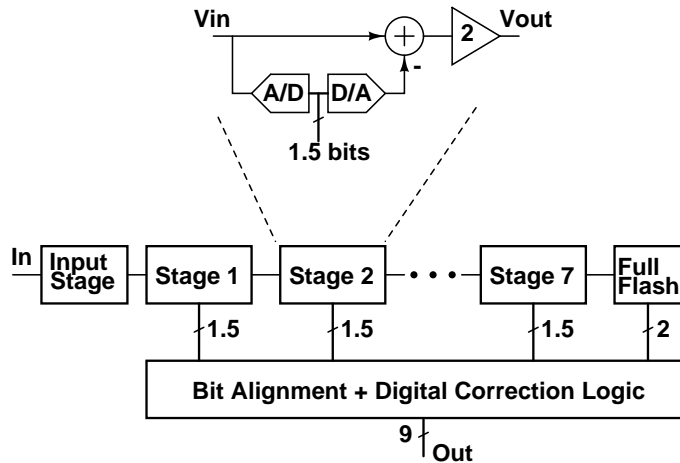
Figure 12.59 Block diagram of the realized ADC.

ing to the simulations, a 14-bit dynamic accuracy is achieved at a 40 MS/s sampling rate. The design presented demonstrates that a current mode track-and-hold circuit can be successfully used for removing glitches from the output of a high-speed current-steering DAC and, as a result, a significant improvement in dynamic performance can be expected.

12.7   1st Switched Opamp Pipelined ADC

12.7.1   Introduction

It has been demonstrated that CMOS ADCs implemented in current mode techniques are capable of operating with supply voltages of 1.5-V and below [240, 241]. Current mode circuits, however, seem to have limited linearity, especially at higher signal frequencies. The switched capacitor (SC) technique, which has an inherently good linearity, has been widely employed in pipelined ADCs operating on supply voltages above 2.5 V, but the insufficient switch overdrive prevents it being used for low-voltage applications in its standard form. The switched opamp technique, which is one of the low-voltage modifications of the SC technique, has, for the first time, been applied to a pipelined ADC in the prototype described in this section and originally published in [8].

12.7.2 ADC Architecture

The 9-bit ADC is realized with the standard 1.5 bits-per-stage pipeline architecture, where the 0.5 bit redundancy in each stage is used for digital correction to relax the requirement for comparator offsets. A block diagram of the ADC is shown in Figure 12.59. Each pipeline stage performs a coarse (in this case three-level) A/D conversion for its input signal and passes the amplified quantization error to the next stage. The quantization error (or residue) is formed by converting the quantization result back to analog form and subtracting it from the input signal. The residue formation and its precise amplification are performed by a multiplying digital-to-analog converter (MDAC).

The operation of the pipeline stage consists of two phases each lasting half a clock cycle. In the first phase the MDAC samples the input signal and the sub-ADC does the A/D conversion. During the second phase the MDAC generates and amplifies the residue, yielding the input signal for the next stage. The successive stages operate in opposite phases and thus the conversion of a sample traverses two stages in a clock cycle.

There is a total of seven stages, like the one whose block diagram is shown in the inset of Figure 12.59. Since the last stage does not need to generate a residue, it is implemented as a 2-bit flash ADC consisting of three comparators and a small number of logic gates. In the reported measurement results the originally 9-bit output is truncated to 8 bits.

12.7.3 MDAC

The operation of the MDAC consists of two phases. During the first phase the input signal is sampled, while the second one is reserved for the subtraction and the amplification. The MDACs in successive stages operate in opposite clock phases.

The MDAC is shown in Figure 12.68. In contrast to its SC counterpart, the feedback capacitors are permanently connected around the amplifier and the input capacitors to the output of the preceding stage. When the MDAC is sampling, the inputs and the outputs of the opamp are connected to V_{DD}. In transition to the amplification they are released, and simultaneously the preceding stage pulls the MDAC inputs to V_{DD}. Consequently, the charge in the input capacitors is transferred to the feedback capacitors. The D/A operation is realized by connecting the $2C$ valued capacitors to the reference voltages according to the bit code produced by the sub A/D converter. The purpose of the C valued capacitors, also controlled by the bit code, is to keep the opamp input common mode level at V_{DD}.
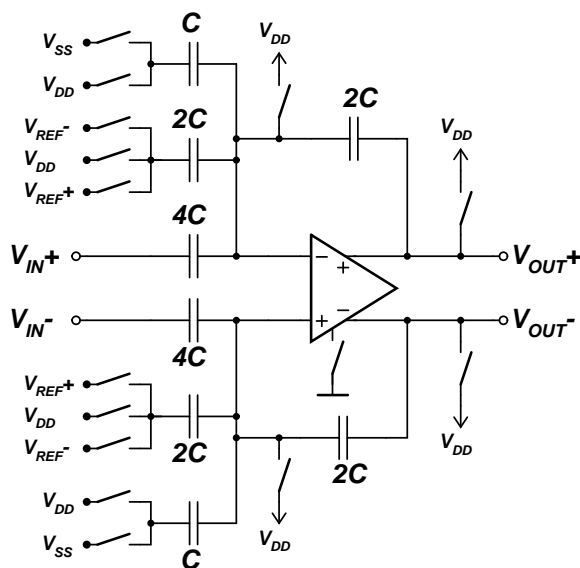
Figure 12.60 The multiplying digital-to-analog converter.

The Miller-type switchable opamp employed in the MDAC has already been described in Section 10.6.2.1.

12.7.4 Comparator

The comparator (Figure 12.61) consists of a preamplifier and a latch. The preamplifier is realized with a differential nMOS pair driving resistor loads. The latch, shown in Figure 12.62, consists of an n-type input pair and a cross-coupled pMOS load with reset switches in parallel. The clock signal controls a switch between the common source node of the input pair and the ground. Although the ADC employs digital correction, careful balancing of the load capacitance in the complementary outputs is needed to guarantee an offset within the tolerable range.

The comparator inputs are connected to the outputs of the previous stage, where the signal common mode level is $V_{DD}/2$. Since this voltage cannot be applied directly to the gate of a transistor, the signal is level-shifted with capacitors. This allows the common mode level at the preamplifier input to be set to V_{DD}. The digital error correction permits rather large error at the comparator decision level, and thus the reference is built into the values of the capacitors C_1 and C_2. The error correction also makes it possible to latch the comparator before the output of the driving stage is fully settled. This is utilized to have the comparison result ready at the time when the
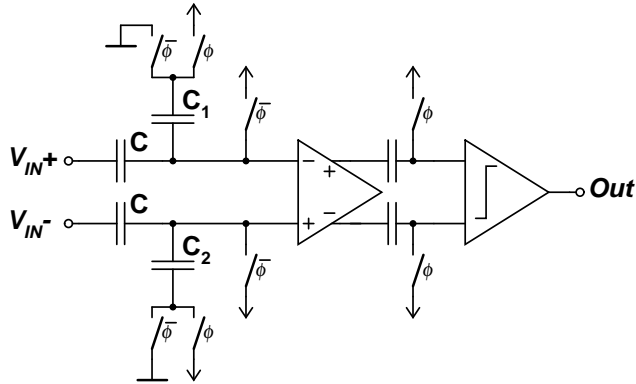
Figure 12.61 Differential comparator.
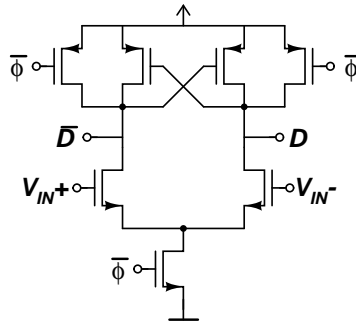
Figure 12.62 Latch of the comparator.

MDAC begins the amplification.

12.7.5 Input Buffer

The active input buffer described in Section 10.7.1 is utilized in this prototype.

12.7.6 Experimental Results

The prototype circuit is fabricated using a 0.5-μm CMOS process with three metal and two polysilicon layers. Its die photograph is shown in Figure 12.63. The total area of the chip is 3.8 mm^2.

Figure 12.64 shows the results of DNL and INL measurements. The DNL errors are within 0.6 LSB. The INL curve obtained is characteristic of all the measured samples. There, the large errors at the edges clearly indicate that the input buffer is not linear enough at the verges of the signal range (1.2 V differential). The INL error in
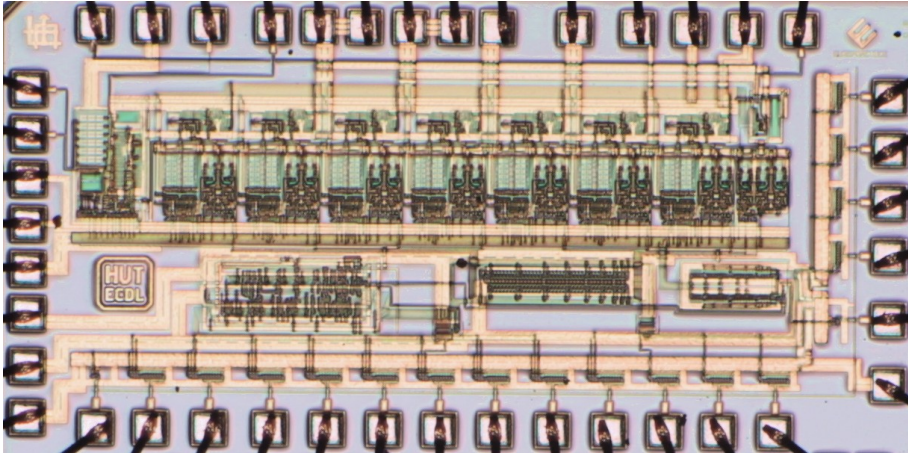
Figure 12.63 Die photograph of 1st switched opamp ADC.

the midrange, however, is very small, suggesting that the A/D itself works well. The INL figures reported in the conference paper [8] were calculated with a code density test program which contained an error. The figures obtained with a corrected version of the program are much better than the incorrect ones reported in the paper.

The SNDR measured as a function of input signal amplitude is shown in Figure 12.65. The measurement was performed using a 200-kHz sinusoidal input signal and an 8-MHz clock. The measured SNDR has a peak value of 44.7 dB, which occurs when the input amplitude is 65% of the full scale. The deterioration of the SNDR with larger amplitudes is explained by the INL curve. The peak SNDR corresponds to 7.1 effective bits.

A spectrum with a 990-kHz 0.6-Vpp input signal is shown in Figure 12.66. The largest spurious component is the third harmonic, which is at the -52-dBc level. The SFDR, which is dominated by the third harmonic, is presented as a function of input frequency in Figure 12.67. It is measured at 5-MHz and 8-MHz clock rates using an input signal whose amplitude is 50% of the full scale. Although the 5-MHz clock gives a roughly 5 dB better SFDR than the 8-MHz clock, the difference in SNDR is much smaller.

The circuit operation was verified with supply voltages ranging from 1.0 to 1.2 volts. All the presented results were measured with a 1.1-V supply. The power consumption at the 8-MHz clock rate with a 200-kHz full scale input signal is 7.8 mW.
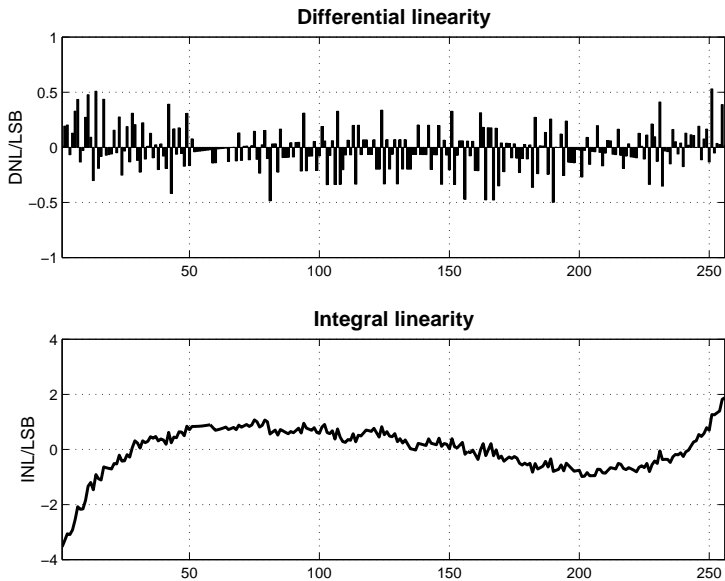
Figure 12.64 Measured static linearity.
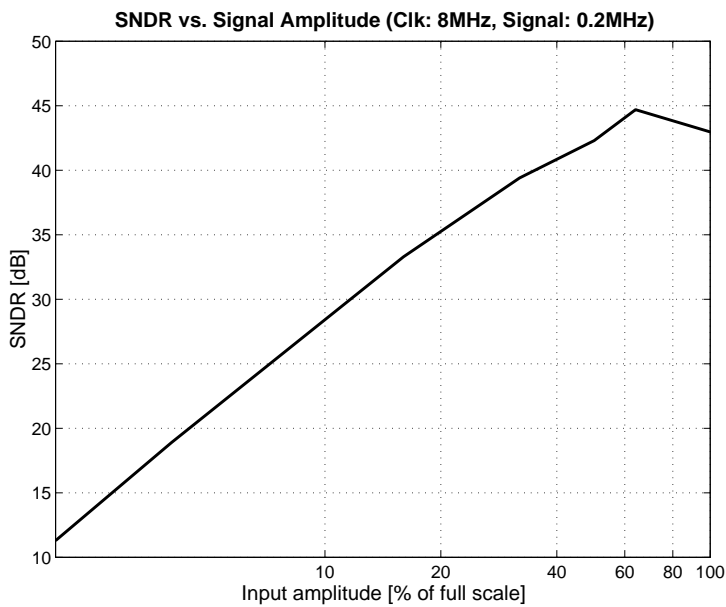
Figure 12.65 SNDR versus signal amplitude.
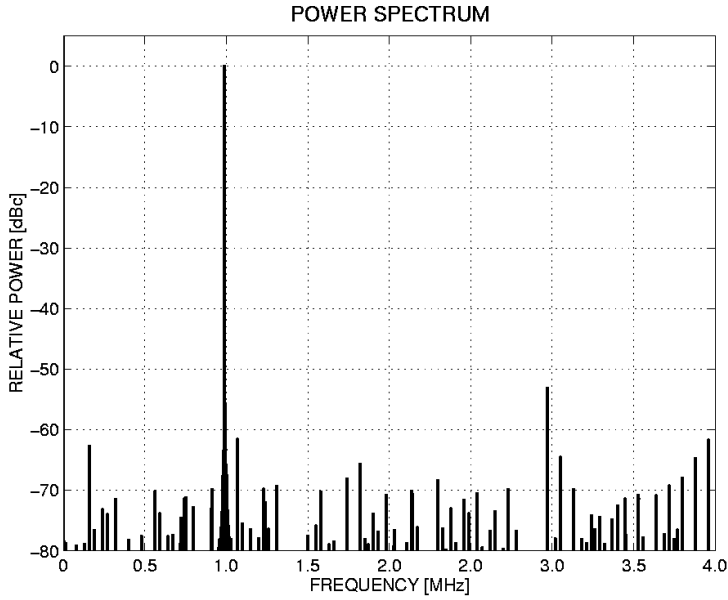
Figure 12.66 Measured spectrum where the signal frequency is 990 kHz and amplitude 50 percent of the full scale.
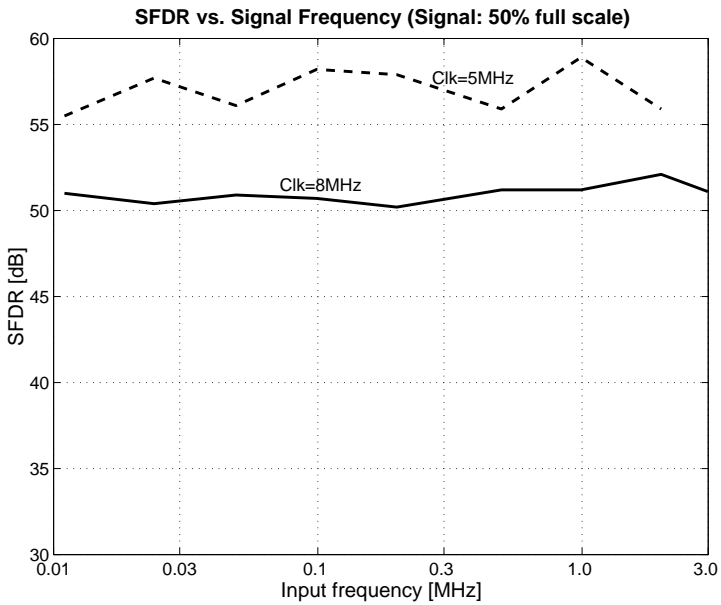
Figure 12.67 Measured SFDR versus signal frequency.
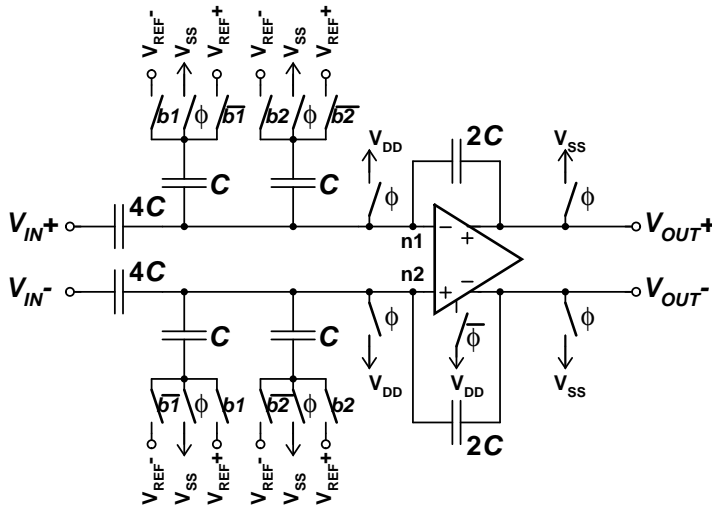
Figure 12.68 Fully differential switched-opamp MDAC.

12.8 2nd Switched Opamp Pipelined ADC

12.8.1 Introduction

This section describes another switched-opamp implementation of a pipelined ADC, which is also published in [3] and [2]. The overall converter architecture is similar to that in the previous section, but most of the circuit blocks have been redesigned to achieve a more robust and less power consuming realization.

12.8.2 MDAC

The developed fully differential SO MDAC is shown in Figure 12.68. In contrast to its SC counterpart, the feedback capacitors are permanently connected around the amplifier and the input capacitors to the output of the preceding stage. Due to this the maximum achievable feedback factor for the MDAC in a 1.5 bits-per-stage architecture is 1/4, while being 1/2 in the SC realization, which makes the SO circuit inherently slower than the SC one.

To understand the circuit's operation let us first look at the DC common mode voltage levels in the circuit. The virtual ground at the opamp input is set to V_{DD}, which is a suitable operating point for an opamp with nMOS input transistors. The voltage level is kept unchanged when switching to the sampling phase by shorting the opamp inputs to V_{DD}. To maximize the voltage swing the signal at the opamp output

is centered in the middle of the supply rails. During the sampling phase the opamp output is in a high impedance state and pulled to V_{SS} by the attached switches. As a result there is a $V_{DD}/2$ change in the voltage level when switching from one phase to the other. Since the MDAC input is connected to the output of the preceding stage, its voltage levels follow a similar pattern, but in the opposite phase.

In the sampling phase—when the clock ϕ is high—the inputs are sampled to the $4C$-valued input capacitors. At the same time the signal voltages on the reference and the feedback capacitors are reset. When entering the amplification phase, the switches at the opamp inputs and outputs are opened. Slightly later, the preceding stage pulls the MDAC inputs to V_{SS}. With the aid of the virtual ground the sampled signal charge in the input capacitors is transferred into the feedback capacitors, resulting in an output voltage that is the input voltage multiplied by the capacitor ratio.

The DAC function is realized with C-valued capacitors, which are connected to either $V_{REF}+$ or $V_{REF}-$ according to the two-bit binary code produced by the stage's sub-ADC. The DAC output, which is added to the sampled voltage, has three possible output values: $+V_{REF}$, $-V_{REF}$, and 0. The first two are achieved by connecting both the capacitors to the same reference voltage, while to get the zero they are connected to the opposite voltages. The advantage of this two-capacitor DAC compared to the DAC in the previous prototype, with a single $2C$-valued capacitor, is the fact that the common mode level for the zero code is the same as for the other codes. This eliminates the extra switched capacitor usually needed in SO circuits to compensate for the common mode level change between the two operating modes.

The minimum supply voltage for the MDAC depends on the gate-source voltage needed to properly turn on the switch transistors and the values of the reference voltages. The reference voltages in turn determine the full-scale signal amplitude, which is equal to the signal swing at the opamp output. This requires the reference levels to be set at least a saturation voltage V_{dsat} apart from the supply rails.

The MDAC uses the switchable opamp introduced in Section 10.6.2.3.

12.8.3 Input Stage

The passive input interface developed for this SO circuit has already been explained in Section 10.7.2. For the sake of convenience it is shown again in Figure 12.69.

The values used for C, Cs, and R are 5 pF, 0.44 pF, and 1.5 kΩ respectively. A total harmonic distortion of 70 dBc is predicted by a simulation made using a 5-MHz clock rate and a 1.7-MHz sinusoidal input signal with a 1.2-Vpp differential amplitude.
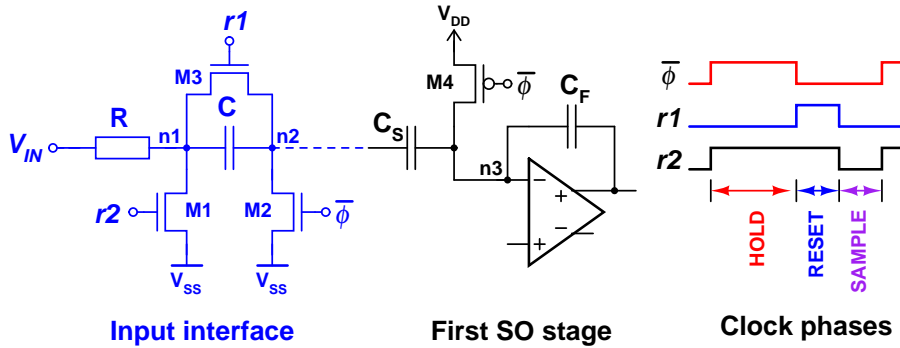
Figure 12.69 A half circuit of the differential input interface and a part of the first pipeline stage.
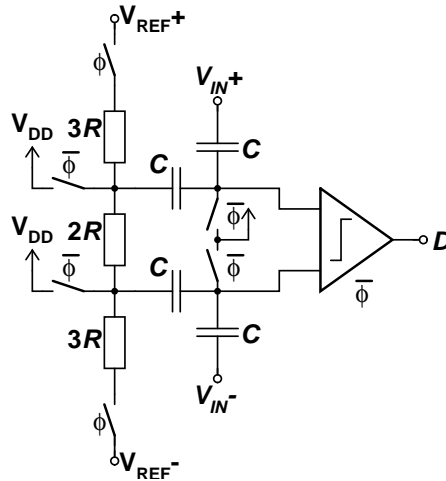
Figure 12.70 Fully differential comparator.

12.8.4 Comparator

In the 1.5-bits-per-stage pipeline architecture there is one redundant quantization level in each sub-ADC which, together with the digital correction, permits $\pm V_{REF}/4$ (± 150 mV in this design) inaccuracy in the comparator decisions. Consequently, the comparator can be a fairly simple dynamic circuit. Since the analog signal path is fully differential, it is desirable that the comparator has a fully differential input, which prevents the use of conventional single-ended comparator topologies where the signal is applied to one input pin and the reference voltage to the other. One possible way to realize a fully differential comparator is to employ a charge summation circuit. The implementation used, e.g. in [104], consists of a latch stage preceded by coupling

capacitors that are precharged to the reference voltages during the reset phase. In the acquisition phase the latch input experiences a voltage that is the sum of the input and the reference; as a result, the comparator decision level is equal to the reference voltage. The implementation used here, shown in Figure 12.70, relies on the same type of architecture.

Due to the switched-opamp implementation of the MDAC, the input capacitors cannot be disconnected from the output of the previous pipeline stage. Thus, another pair of capacitors is needed for adding the reference voltages. During the reset phase the previous converter stage pulls the input capacitors to V_{SS} and the inputs of the latch are reset to V_{DD}. At the same time the reference capacitors are connected to V_{DD}. When the driving pipeline stage starts its amplification phase, the reset switches of the comparator are released and the resistor string, to which the reference capacitors are attached, is connected between the global positive and negative reference voltages. The resistor string, which is common for all the comparators in the same pipeline stage, provides the voltages $+V_{REF}/4$ and $-V_{REF}/4$, setting the comparator decision level accordingly.

There are two reasons why the quarter reference is realized with a resistor chain instead of simply scaling the reference capacitor values. First, when using equal size input and reference capacitors, the common mode voltage level in the latch input is automatically set to V_{DD}. Otherwise, an extra pair of capacitors would have been needed to correct the voltage level. The second reason is the desire to keep the capacitors as small as possible. The capacitor value $C/4$ would have been too small to implement without increasing the absolute value of the unit capacitor C.

The dynamic latch is similar to the one used in the first version and shown in Figure 12.62. The digital error correction makes it possible to trigger the comparator before the output of the driving stage is fully settled. This is utilized to make sure that the A/D conversion result is ready at the time when the MDAC begins the amplification. The comparators are synchronized with the MDAC by adding a digital latch stage after them.

12.8.5   Experimental Results

The prototype chip is implemented in a 0.5-μm triple-metal double-poly CMOS technology with 610-mV Vth for both nMOS and pMOS transistors. A photograph of the die, witch has an active area of 1.3 mm^2, is shown in Figure 12.71. The prototype is packaged in a 44-pin CLCC package and the measurements are performed using a test board made of 2-layer PCB and having a socket for the chip.
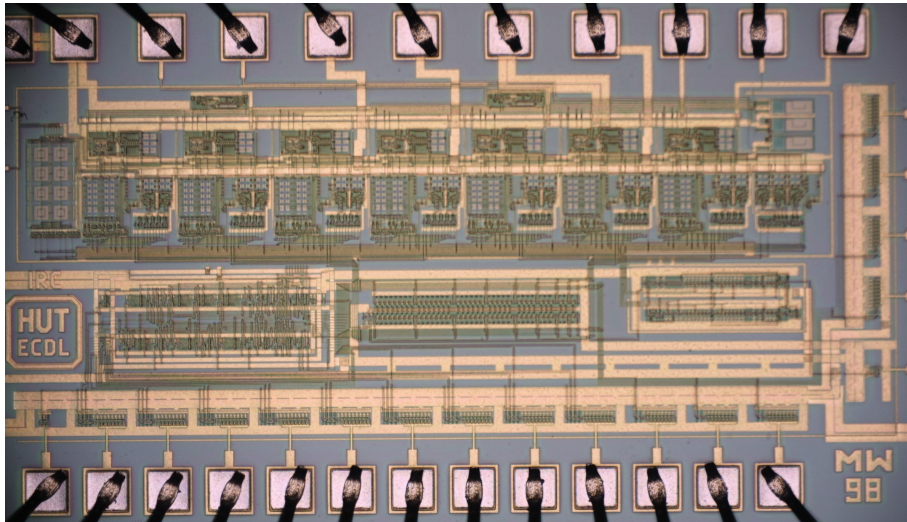
Figure 12.71 Photograph of the prototype chip.

The operation of the circuit is verified with a supply voltage range from 0.95 V to 1.6 V. All the results given were obtained with a 1.0-V supply unless otherwise specified. The reference voltages are set 600 mV apart and symmetrically between the supply rails, resulting a ±600-mV differential input signal range.

The measured DNL and INL curves are presented in Figure 12.72. For nine measured samples the maximum DNL and INL are 0.6 LSB and 1.1 LSB respectively. The SNDR versus the input signal amplitude measured at 1.0-V and 1.5-V supply voltages and 5-MHz and 14-MHz respective clock rates is shown in Figure 12.73. In both cases the peak SNDR is obtained with the full-scale signal amplitude and has a value of 50.0 dB, which corresponds to 8.0 effective bits.

Figure 12.74 shows a spectrum where a 1.5-MHz full-scale sine wave is sampled at 5.0 MS/s, the supply voltage being 1.0 V. A 14-MS/s spectrum obtained with the 1.5-V supply is presented in Figure 12.75. With the low supply voltage the spurious free dynamic range is limited by the second harmonic and it is 59.5 dB. The level of the harmonic is raised as the bias current is increased along with the supply voltage, reducing the SFDR to 55.3 dB at 1.5 V.

The ADC performance is limited by the static nonlinearities, since the SNDR is virtually constant up to a 6-MHz clock frequency, after which there is a sudden collapse in the signal quality, indicating timing problems in the digital circuitry. When the supply voltage is increased to 1.5 volts the same phenomenon is observed at the 14.5-MHz clock frequency. The power consumption from a 1.0-V supply at the 5.0-
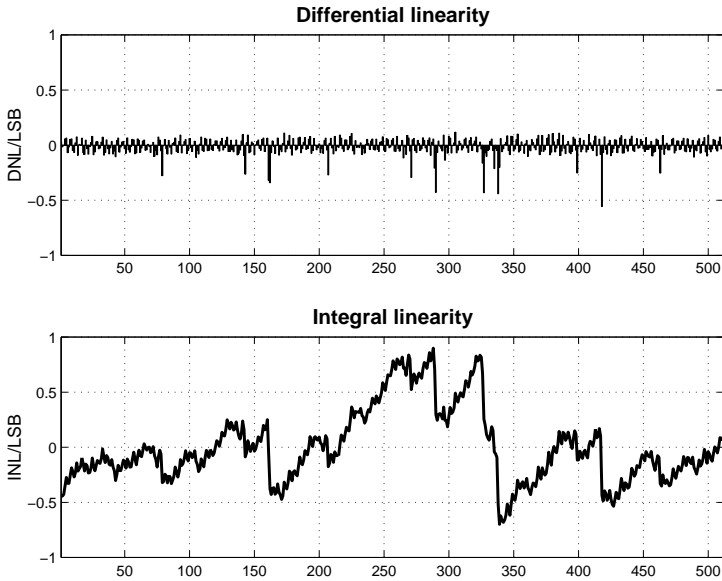
Figure 12.72 Measured DNL and INL.
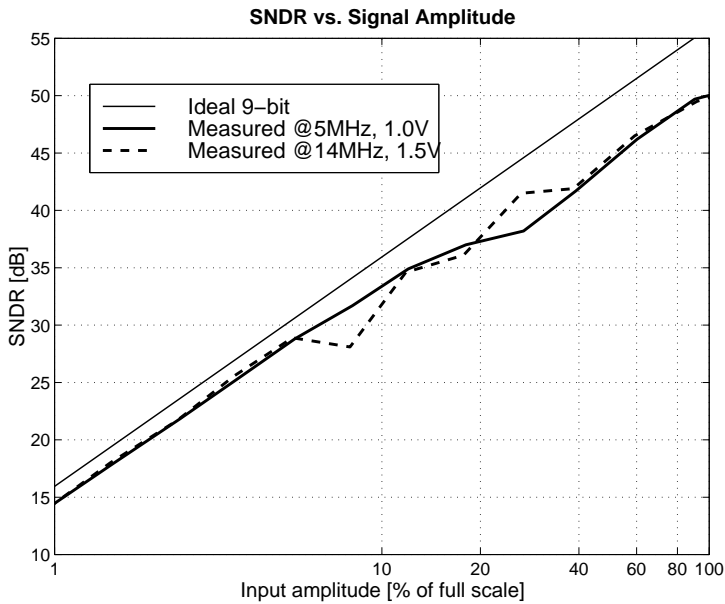
Figure 12.73 SNDR versus signal amplitude. At the 1.0-V supply voltage the clock rate is 5 MHz and the signal frequency 200 kHz. The 1.5-V curve is obtained with a 14-MHz clock and a 1.5-MHz signal.
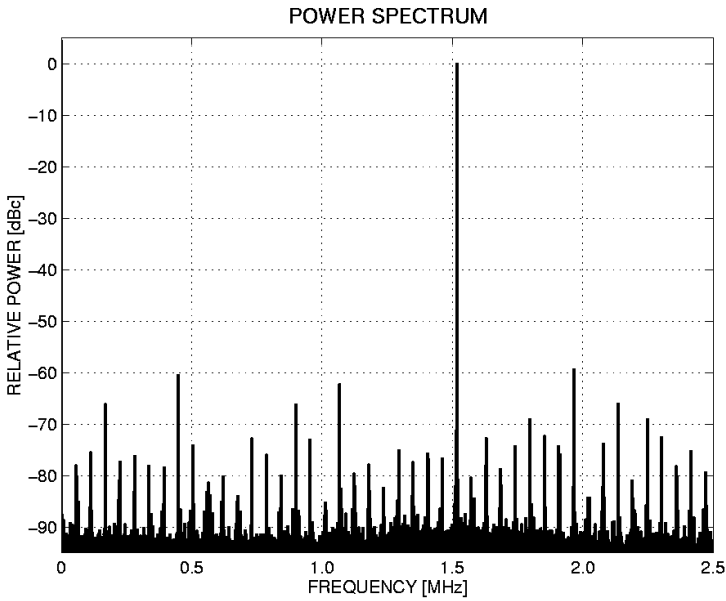
Figure 12.74 A spectrum measured at a 5-MHz clock rate using a 1.0-V supply voltage. The signal is a 1.5-MHz full-scale sine wave.
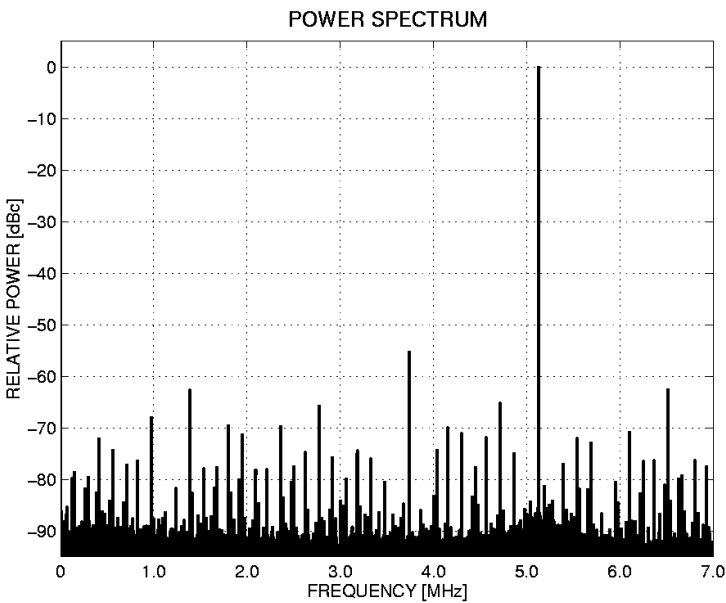
Figure 12.75 A spectrum measured at a 14-MHz clock rate using a 1.5-V supply voltage. The signal is a 5.1-MHz full-scale sine wave.

Table 12.4 Summary of ADC performance.

Supply voltage	1.0 V	1.5 V
Resolution	9 bits	
Conversion rate	5 MS/s	14 MS/s
DNL	0.6 LSB	0.6 LSB
INL	1.1 LSB	1.1 LSB
SNDR	50.0 dB	50.0 dB
SFDR	59.5 dB	55.3 dB
Input range	± 600 mV	± 600 mV
Power consumption	1.6 mW	8.2 mW
Technology	0.5-μm CMOS	
Active area	1.3 mm^2	

MS/s sampling rate is 1.6 mW. In addition to increasing the supply voltage to 1.5 V, the bias current has to be tripled to achieve the 14-MS/s sampling rate. This increases the power consumption to 8.2 mW, which is still very low compared to the 36 mW achieved in [104] with the same supply voltage and sampling rate. The measured performance is summarized in Table 12.4.

The measurements demonstrate that the SO technique, previously considered an immature and performance-limited technique, can be used to realize low-voltage ADCs and provides a performance well-matched to that found with other low-voltage realizations.

Conclusions

Technology scaling will bring advantages to SC circuits for the next few technology generations, but after that the effect of the decreasing supply voltage on the signal-to-noise ratio will start to dominate over the positive effects of shrinking transistor dimensions. Even before that, maximizing the signal range is essential for exploiting the benefits of technology scaling. The signal range has the largest effect on the opamps, making the use of a rail-to-rail output stage mandatory. The switches do not have major difficulties in adapting to the nominal supply voltage of the technology. On the contrary, when a smaller-than-nominal supply voltage is used or a performance increase is pursued, techniques such as gate voltage bootstrapping or the switched-opamp technique are needed to guarantee a small enough switch on-resistance.

In this work the utilization of the switched-opamp technique in pipelined ADCs has been demonstrated. Connecting the SO circuit to the outside world has been solved with a passive input interface circuit. The main limitation of the SO technique is the low speed caused by opamp switching and the decreased feedback factor. Thus, probably a better approach for low-voltage circuits requiring high speed is the selective use of bootstrapped switches and the utilization of different common mode signal levels at the opamp input and the output. The applications of the SO technique are in the areas where speed is not the most critical parameter.

In wide-band radio receivers, moving the signal digitization into a high interme-diate frequency is attractive. It has been demonstrated that the sampling can be done with relatively high linearity by using bootstrapped switches. A more fundamental problem is jitter, which can be alleviated only by extensive oversampling.

The elimination of the bulk effect and signal-dependent drain and source junction capacitances is essential in a highly linear bootstrapped switch. This can most easily be accomplished with a triple-well process. The same, although with a smaller linearity boost, has here been demonstrated with a standard CMOS technology.

Time interleaving is a way to extend the conversion rates of ADCs beyond the

technology-determined limits of one stand-alone A/D channel. For eliminating the non-uniform sampling effects caused by timing skew between the parallel channels, a front-end S/H circuit is practically mandatory. As a result, it becomes a speed bottle-neck, typically limiting the number of parallel channels to a maximum of four.

Moore's law can be most effectively exploited by doing things digitally. This can be realized on the system level by moving analog functions into the digital domain, but also inside the analog system blocks by using even extensive amounts of digital circuitry to correct and compensate for the imperfections of the analog circuitry. In ADCs this means incorporating logic and memory for calibrating and correcting the offsets and the component mismatch.

Bibliography

[1] M. Waltari, L. Sumanen, T. Korhonen, K. Halonen, "A Self-Calibrated Pipeline ADC with 200MHz IF-sampling Frontend," *2002 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pp. 314–315, Feb. 2002. 2, 3, 110, 192

[2] M. Waltari, K. Halonen, "1-V 9-Bit Pipelined Switched-Opamp ADC," *IEEE J. Solid-State Circuits*, vol. 36, pp. 129–134, Jan. 2001. 3, 4, 58, 134, 138, 148, 155, 231

[3] M. Waltari, K. Halonen, "1.0-Volt, 9-bit Pipelined CMOS ADC," in *proc 26th European Solid-State Circuits Conference*, pp. 360–363, Sep. 2000. 231

[4] L. Sumanen, M. Waltari, K. Halonen, "A 10-bit 200 MS/s CMOS Parallel Pipeline A/D Converter," *IEEE J. Solid-State Circuits*, vol. 36, pp. 1048–1055, Jul. 2001. 3, 53, 57, 98, 121, 180, 181

[5] L. Sumanen, M. Waltari, K. Halonen, "A 10-bit 200 MS/s CMOS Parallel Pipeline A/D Converter," in *proc 26th European Solid-State Circuits Conference*, pp. 440–443, Sep. 2000. 180

[6] M. Waltari, K. Halonen, "Fully differential switched opamp with enhanced common mode feedback," *IEE Electronics Letters*, vol. 34, no. 23, pp. 2181–2182, Nov. 12, 1998. 3, 145

[7] M. Waltari, K. Halonen, "A Switched-Opamp With Fast Common Mode Feedback," in *proc IEEE International Conference on Electronics, Circuits and Systems*, vol. III, pp. 1523–1525, Sep 1999.

[8] M. Waltari, K. Halonen, "An 8-bit Low-Voltage Pipelined ADC Utilizing Switched-Opamp Technique," in *Proceedings of the 25th European Solid-State Circuits Conference*, pp. 174–177, Sep. 1999. 3, 4, 58, 134, 144, 153, 224, 228

[9] M. Waltari, K. Halonen, "Timing Skew Insensitive Switching for Double-Sampled Circuits," in proc *IEEE International Symposium on Circuits and Systems*, vol. II, pp. 61–64, May 1999. 3, 121, 130, 176

[10] M. Waltari, K. Halonen, "A 220-MSample/s CMOS Sample-and-Hold Circuit Using Double-Sampling," *Analog Integrated Circuits and Signal Processing*, vol. 18, pp. 21–31, Jan. 1999. 3, 79, 81, 121, 169

[11] M. Waltari, K. Halonen, "A 10-bit 220-MSample/s CMOS Sample-and-Hold Circuit," in proc. *IEEE International Symposium on Circuits and Systems*, vol. I, pp. 253–256, May 1998. 3, 121, 169

[12] M. Waltari, K. Halonen, "A 10-bit, 130-Msample/s CMOS Sample-and-Hold circuit," in proc. *NORCHIP'97*, pp. 31–35, 1997. 169

[13] M. Waltari, K. Halonen, "A 10-bit 130-MSample/s CMOS Sample-and-Hold Circuit," *HUT IRC Radio Communications Research Workshop*, 1997.

[14] M. Waltari, K. Halonen, "Reference Voltage Driver for Low-Voltage CMOS A/D Converters," in *proc IEEE International Conference on Electronics, Circuits and Systems*, pp. 28–31, Dec. 2000. 3, 164

[15] M. Waltari, J. Pirkkalaniemi, L. Sumanen, M. Kosunen, K. Halonen, "A 14-Bit, 40-MS/s DAC with Current Mode Deglitcher," in proc. *NORCHIP'01*, pp. 210–215, Nov. 2001. 214

[16] J. Pirkkalaniemi, M. Waltari, L. Sumanen, M. Kosunen, K. Halonen, "A 14-Bit, 40-MS/s DAC with Current Mode Deglitcher," accepted in ISCAS'02, May 2002. 4, 214

[17] M. Waltari, *Integration of High-Speed CMOS Sample-and-Hold Circuits*, Licentiate's Thesis, Helsinki University of Technology, 1999. 2

[18] *International Technology Roadmap for Semiconductors*, Semiconductor Industry Association, 1999. 5

[19] K. Bult, "Analog Design in Deep Sub-Micron CMOS," in *Proceedings of the 26th European Solid-State Circuits Conference*, pp. 11–17, Sep. 2000. 5

[20] R. Castello, P. R. Gray, "Performance Limitations in Switched-Capacitor Filters," *IEEE Trans. Circuits and Systems,* vol. CAS-32, pp. 865–876, Sep. 1985. 6

[21] G. Groenewold, "Optimal Dynamic Range Integrators," *IEEE Trans. Circuits and Systems–I,* vol. 39, pp. 614–627, Aug. 1992. 6

[22] A-J. Annema, "Analog Circuit Performance and Process Scaling," *IEEE Trans. Circuits and Systems–II,* vol. 46, pp. 711–725, Jun. 1999. 6, 12

[23] F. Maloberti, F. Francesconi, P. Malcovati, O. J. A. P. Nys, "Design Considerations on Low-Voltage Low-Power Data Converters," *IEEE Trans. Circuits and Systems–I,* vol. 42, pp. 653–863, Nov. 1995. 6, 82, 83

[24] R. Castello, F. Montecchi, F. Rezzi, A. Baschirotto, "Low-VoltageAnalog Filters," *IEEE Trans. Circuits and Systems–I,* vol. 42, pp. 827–840, Nov. 1995. 6

[25] A. Matsuzawa, "Low-Voltage and Low-Power Circuit Design for Mixed Analog/Digital Systems in Portable Equipment," *IEEE J. Solid-State Circuits*, vol. 29, pp. 470–480, Apr. 1994. 6

[26] B. J. Hosticka, W. Brockherde, D. Hammerschmidt, R. Kokozinski, "Low-Voltage CMOS Analog Circuits," *IEEE Trans. Circuits and Systems–I,* vol. 42, pp. 864–872, Nov. 1995. 6

[27] E. A. Vittoz, "Low-Power Design: WAys to Approach the Limits," in *Dig. Tech. Papers International Solid-State Circuits Conference*, pp. 14–18, Feb. 1994. 7

[28] *SOI CMOS Technology, devices and circuits for low-power low-voltage analog and microwave applications*, Euro Practice Analog Training Course, 1999. 10

[29] K-Y. Toh, P-K. Ko, R. G. Meyer, "An Engineering Model for Short-Channel MOS Devices," *IEEE J. Solid-State Circuits*, vol. 23, pp. 950–958, Aug. 1988. 12

[30] W. Sansen, "Analog Circuit Design inScaled CMOS Technology," in *Symposium on VLSI Circuits Dig. of Tech. Papers*, pp. 8–11, 1996. 12

[31] W. Sansen, M. Steyaert, V. Peluso, E. Peeters, "Toward Sub 1V Analog Integrated Circuits in Submicron Standard CMOS Technologies," in *Dig. Tech. Papers International Solid-State Circuits Conference*, pp. 186–187, Feb. 1998. 12

[32] K. B. Lakshmikumar, R. A. Hadaway, M. A. Copeland, "Characterization and Modelling of Mismatch in MOS Transistors for Precision Analog Design," *IEEE J. Solid-State Circuits*, vol. SC-21, pp. 1057–1066, Dec. 1986. 13

[33] M. J. Pelgrom, A. C. J. Duinmaijer, A. P. Welbers, "Matching Properties of MOS Transistors," *IEEE J. Solid-State Circuits*, vol. 24, pp. 1433–1440, Oct. 1989.   13

[34] M.Steyaert, K. Uyttenhove, "Speed-Power-Accuracy Trade-off In high-speed Analog-to-digital converters: Now and in the future ...," in *Workshop on Advances in Analog Circuit Design*, April 2000   14

[35] M. Steyaert, V. Peluso, J. Bastos, P. Kinget, W. Sansen, "Custom Analog Low Power Design: The problem of low voltage and mismatch," in *Proc. IEEE Custom Integrated Circuits Conference*, pp. 285–292, 1997.   14

[36] J. Crols, M. Steyaert, "Switched-opamp: an approach to realize full CMOS switched capacitor circuits at very low power supply voltages," *IEEE J. Solid-State Circuits*, vol. 29, pp. 936–924, Aug. 1994.   18, 134, 141, 152

[37] R. van de Plassche, *Integrated Analog-to-Digital and Digital-to-Analog Converters*, Kluwer Academic Publishers, 1994.   21, 34, 77, 166

[38] B. Razavi, *Principles of Data Conversion System Design*, IEEE Press, 1995. 21

[39] S. Lindfors, *CMOS Baseband Integrated Circuit Techniques for Radio Receivers*, Doctoral Thesis, Helsinki University of Technology, Jul. 2000.   24

[40] M. Sauerwald, "Effects of Aperture Time and Jitter in a Sampled Data System," National Semiconductor Corporation, Application note AD-03, 1994.   27

[41] M. Shinagawa, Y. Akazawa, T. Wakimoto, "Jitter Analysis of High-Speed Sampling Systems," *IEEE J. Solid-State Circuits*, vol. 25, pp. 220–224, Feb. 1990. 28

[42] R. Gregorian, G. C. Temes, *Analog MOS Integrated Circuits for Signal Processing*, John Wiley & Sons, Inc. 1986.   29

[43] K. R. Stafford, P. R. Gray, R. A. Blanchard, "A Complete Monolithic Sample/Hold Amplifier," *IEEE J. Solid-State Circuits*, vol. SC-9, pp. 381–387, Dec. 1974.   30

[44] M. Nayebi, B. A. Wooley, "A 10-bit Video BiCMOS Track-and-Hold Amplifier," *IEEE J. Solid-State Circuits*, vol. 24, pp. 1507–1516, Dec. 1989.   30

[45] K. Kattmann, J. Barrow, "A Technique for Reducing Differential Nonlinearity Errors in Flash A/D Converters," in *Dig. Tech. Papers International Solid-State Circuits Conference*, pp. 170–171, Feb. 1991. 37

[46] K. Bult, A. Buchwald, "An Embedded 240-mW 10-b 50-MS/s CMOS ADC in 1mm^2," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1887–1895, Dec. 1997. 37, 42

[47] H. Kimura, A. Matsuzawa, T. Nakamura, S. Sewada, "A 10-b 300-MHz Interpolated-Parallel A/D Converter," *IEEE J. Solid-State Circuits*, vol. 28, pp. 438–446, Apr. 1993. 37

[48] M. Choi, A. A. Abidi, "A 6b 1.3GSample/s A/D Converter in 0.35μm CMOS," in *Dig. Tech. Papers International Solid-State Circuits Conference*, pp. 126–127, Feb. 2001. 37

[49] G. Geelen, "A 6b 1.1GSample/s CMOS A/D Converter," in *Dig. Tech. Papers International Solid-State Circuits Conference*, pp. 128–129, Feb. 2001. 37

[50] A. G. Dingwall, V. Zazzu, "An 8-MHz CMOS subranging 8-Bit A/D converter," *IEEE J. Solid-State Circuits*, vol. SC-20, pp. 1138–1143, Dec. 1985. 37

[51] R. C. Taft, M. R. Tursi, "A 100-MS/s 8-b CMOS Subranging ADC with Sustained Parametric Performance from 3.8 V Down to 2.2 V," *IEEE J. Solid-State Circuits*, vol. 36, pp. 331–338, Mar. 2001. 39

[52] H. van der Ploeg, G. Hoogzaad, H. A. H. Termeer, M. Vertregt, R. L. J. Rooves, "A 2.5V 12b 54MSample/s 0.25μm CMOS ADC in 1mm^2," in *Dig. Tech. Papers International Solid-State Circuits Conference*, pp. 132–133, Feb. 2001. 39

[53] B. Nauta, A. G. W. Venes, "A 70-MS/s 110-mW 8-b CMOS Folding and Interpolating A/D Converter," *IEEE J. Solid-State Circuits*, vol. 30, pp. 1302–1308, Dec. 1995. 41

[54] A. G. W. Venes, R. J. van de Plassche, "An 80-MHz, 80-mW 8-b CMOS Folding A/D Converter with Distributed Track-and-Hold Preprocessing," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1846–1853, Dec. 1996. 42

[55] P. Vorenkamp, R. Roovers, "A 12-b, 60-MSample/s Cascaded Folding and Interpolating ADC," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1876–1886, Dec. 1997. 42

[56] M.-J. Choe, B.-S. Song, K. Bacrania, "A 13-b 40-MSample/s CMOS Pipelined Folding ADC with Background Offset Trimming," *IEEE J. Solid-State Circuits*, vol. 35, pp. 1781–1790, Dec. 2000. 42

[57] Y.-T. Wang, B. Razavi, "An 8-Bit 150-MHz CMOS A/D Converter," *IEEE J. Solid-State Circuits*, vol. 35, pp. 308–317, Mar. 2000. 42

[58] M. P. Flynn, B. Sheahan, "A 400-Msample/s. 6-b CMOS Folding and Interpolating ADC," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1132–1938, Dec. 1998. 42

[59] S. H. Lewis, P. R. Gray, "A Pipelined 5-Msample/s 9-bit Analog-to-Digital Converter," *IEEE J. Solid-State Circuits*, vol. sc-22, pp. 954–961, Dec. 1987. 46, 47, 68, 69

[60] D. Macq, P. G. A. Jespers, "A 10-Bit Pipelined Switched-Current A/D Converter," *IEEE J. Solid-State Circuits*, vol. 29, pp. 967–971, Aug. 1994. 46

[61] C.-Y. Wu, C.-C. Chen, J.-J. Cho, "A CMOS Transitor-Only 8-b 4.5-Ms/s Pipelined Analog-to-Digital Conveter Using Fully-Differential Current-Mode Circuit Techniques," *IEEE J. Solid-State Circuits*, vol. 30, pp. 522–532, May 1995. 46

[62] B. Ginetti, P. Jespers, "A 1.5 MS/s 8-bit Pipelined RSD A/D Converter," *ESSCIRC Dig. Tech. Papers*, pp. 137–140, Sep. 1990. 47

[63] B. Ginetti, P. G. A. Jespers, A. Vandemeulebroecks, "A CMOS 13-b cyclic RSD A/D converter," *IEEE J. Solid-State Circuits*, vol. 27, pp. 957–965, Jul. 1992. 47

[64] S. H. Lewis, H. S. Fetterman, G. F. Gross, Jr., R. Ramachandran, T. R. Viswanathan, "A 10-b 20-Msample/s Analog-to-Digital Converter," *IEEE J. Solid-State Circuits*, vol. 27, pp. 351–358, Mar. 1992. 47

[65] E. G. Soenen, R. L. Geiger, "An Architecture and An Algorithm for Fully Digital Correction of Monolithic Pipelined ADC's," *IEEE Trans. Circuits and Systems–II,* vol. 42, pp. 143–153, Mar. 1995. 47, 54

[66] S. Sutarja, P. Gray, "A Pipelined 13-bit, 250-ks/s, 5-V Analog-to-Digital Converter," *IEEE J. Solid-State Circuits*, vol. 23, pp. 1316–1323, Dec. 1988. 49

[67] P. C. Yu, H.-S. Lee, "A Pipelined A/D Conversion Technique with Near-Inherent Monotonicity," *IEEE Trans. Circuits and Systems–II,* vol. 42, pp. 500–502, Jul. 1995. 49

[68] P. C. Yu, H.-S. Lee, "A 2.5-V, 12-b, 5-MSample/s Pipelined CMOS ADC," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1854–1861, Dec. 1996. 49

[69] P. W. Li, M. J. Chin, P. R. Gray, R. Castello, "A Ratio-Independent Algorithmic Analog-to-Digital Conversion Technique," *IEEE J. Solid-State Circuits*, vol. SC-19, pp. 828–836, Dec. 1984. 49

[70] B.-S. Song, M. F. Tompsett, K. R. Lakshmikumar, "A 12-bit 1-Msample/s Capacitor Error-Avaraging Pipelined A/D Converter," *IEEE J. Solid-State Circuits*, vol. 23, pp. 1324–1333, Dec. 1988. 49, 102

[71] H.-S, Chen, B.-S. Song, K. Bacrania, "A 14-b 20-MSamples/s CMOS Pipelined ADC," *IEEE J. Solid-State Circuits*, vol. 36, pp. 997–1001, Jun. 2001. 49

[72] Y. Chiu, "Inherently Linear Capacitor Error-Avaraging Techniques for Pipelined A/D Conversion," *IEEE Trans. Circuits and Systems–II,* vol. 47, pp. 229–232, Mar. 2000. 49

[73] T. B. Cho, P.R. Gray, "A 10 b, 20 Msample/s, 35 mW Pipeline A/D Converter," *IEEE J. Solid-State Circuits*, vol. 30, pp. 166–172, Mar. 1995. 50, 82, 83, 103, 184

[74] D. W. Cline, P. R. Gray, "A Power Optimized 13-b 5 Msamples/s Pipelined Analog-to-Digital Converter in 1.2 μm CMOS," *IEEE J. Solid-State Circuits*, vol. 31, pp. 294–303, Mar. 1996. 50, 54

[75] W. Yang, D. Kelly, I. Mehr, M. T. Sayuk, L. Singer, "A 3-V 340-mW 14-b 75-Msample/s CMOS ADC With 85-dB SFDR at Nyquist Input," *IEEE J. Solid-State Circuits*, vol. 36, pp. 1931–1936, Dec. 2001. 51, 54

[76] S. H. Lewis, "Optimizing the Stage Resolution in Pipelined, Multistage, Analog-to-Digital Converters for Video-Rate Applications," *IEEE Trans. Circuits and Systems–II,* vol. 39, pp. 516–523, Aug. 1992. 51, 52

[77] P. T. F. Kwok, H. C. Luong, "Power Optimization for Pipeline Analog-to-Digital Converters," *IEEE Trans. Circuits and Systems–II,* vol. 46, pp. 549–553, May 1999. 54

[78] L. A. Singer, T. L. Brooks, "A 14-Bit 10-MHz Calibration-Free CMOS Pipelined A/D Converter," *Symposium on VLSI Circuits Digest of Technical Papers*, pp. 94–95, 1996. 54

[79] S.-U. Kwak, B.-S. Song, K. Bacrania, "A 15-b, 5-Msample/s Low-Spurious CMOS ADC," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1866–1875, Dec. 1997. 54, 56

[80] H.-S. Lee, D. A. Hodges, P. R. Gray, "A Self-Calibrating 15 Bit CMOS A/D Converter," *IEEE J. Solid-State Circuits*, vol. SC-19, pp. 813–819, Dec. 1984. 54

[81] Y.-M. Lin, B. Kim, P. R. Gray, "A 13-b 2.5-MHz Self-Calibrated Pipelined A/D Converter in 3-μm CMOS," *IEEE J. Solid-State Circuits*, vol. 26, pp. 628–635, Apr. 1991. 54, 68

[82] S.-H. Lee, B.-S. Song, "Digital-Domain Calibration for Multistep Analog-to-Digital Converters," *IEEE J. Solid-State Circuits*, vol. 27, pp. 1679–1688, Dec. 1992. 54, 199

[83] A. N. Karanicolas, H.-S. Lee, K. L. Bacrania, "A 15-b 1-Msample/s Digitally Self-Calibrated Pipeline ADC," *IEEE J. Solid-State Circuits*, vol. 28, pp. 1207–1215, Dec. 1993. 54

[84] M. K. Mayes, S. W. Chin, "A 200 mW, 1 Msample/s, 16-b Pipelined A/D Converter with On-Chip 32-b Microcontroller," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1862–1872, Dec. 1996. 54

[85] G.-C. Ahn, H.-C. Choi, S.-I. Lim, S.-H. Lee, C.-D. Lee, "A 12-b, 10-MHz, 250-mW CMOS A/D Converter," *IEEE J. Solid-State Circuits*, vol. 31, pp. 2030–2035, Dec. 1996. 54

[86] I. E. Opris, L. D. Lewicki, B. C. Wong, "A Single-Ended 12-bit 20 Msample/s Self-Calibrating Pipeline A/D Converter," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1898–1903, Dec. 1998. 54

[87] J. M. Ingino, B. A. Wooley, "A Contunuously Calibrated 12-b, 10-MS/s, 3.3-V A/D Converter," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1920–1931, Dec. 1998. 56

[88] U.-K. Moon, B.-S. Song, "Background Digital Calibration Techniques for Pipelined ADC's," *IEEE Trans. Circuits and Systems–II,* vol. 44, pp. 102–109, Feb. 1997. 56

[89] O. E. Erdogan, P. J. Hurst, S. H. Lewis, "A 12-b Digital-Background-Calibrated Algorithmic ADC with -90-dB THD," *IEEE J. Solid-State Circuits*, vol. 34, pp. 1812–1820, Dec. 1999. 56

[90] T.-H. Shu, B.-S. Song, K. Bacrania, "A 13-b 10-Msample/s ADC Digitally Calibrated with Oversampling Delta-Sigma Converter," *IEEE J. Solid-State Circuits*, vol. 30, pp. 443–452, Apr. 1995. 56

[91] I. Galton, "Digital Cancellation of D/A Converter Noise in Pipelined A/D Converters," *IEEE Trans. Circuits and Systems–II,* vol. 47, pp. 185–196, Mar. 2000. 56

[92] W. C. Black, Jr., D. A. Hodges, "Time Interleaved Converter Arrays," *IEEE J. Solid-State Circuits*, vol. SC-15, pp. 1022–1029, Dec. 1980. 56, 57, 180

[93] C. S. G. Conroy, D. W. Cline, P. R. Gray, "An 8-b 85-MS/s Parallel Pipeline A/D Converter in 1-μm CMOS," *IEEE J. Solid-State Circuits*, vol. 28, pp. 447–454, Apr. 1993. 56, 57, 119, 180, 186

[94] M. Yotsuyanagi, T. Etoh, K. Hirata, "A 10-b 50-MHz Pipelined CMOS A/D Converter with S/H," *IEEE J. Solid-State Circuits*, vol. 28, pp. 292–300, Mar. 1993. 56, 57, 106

[95] K. Nakamura, M. Hotta, L. R. Carley, D. J. Allstot, "An 85 mW, 10 b, 40 Msample/s CMOS Parallel-Pipelined ADC," *IEEE J. Solid-State Circuits*, vol. 30, pp. 173–183, Mar. 1995. 56, 186

[96] K. Nagaraj, H. S. Fetterman, J. Anidjar, S. H. Lewis, R. G. Renninger, "A 250-mW, 8-b, 52-Msamples/s Parallel-Pipelined A/D Converter with Reduced Number of Amplifiers," *IEEE J. Solid-State Circuits*, vol. 32, pp. 312–320, Mar. 1997. 56, 123

[97] W. Bright, "8b 75MSample/s 70mW Parallel Pipelined ADC Incorporating Double Sampling," *ISSCC Dig. Tech. Papers*, pp. 146–147, Feb. 1998. 56, 121, 180

[98] K. Y. Kim, N. Kusayanagi, A. A. Abidi, "A 10-b, 100-MS/s CMOS A/D Converter," *IEEE J. Solid-State Circuits*, vol. 32, pp. 302–311, Mar. 1997. 56, 57, 69, 70, 180

[99] K. C. Dyer, D. Fu, S. H. Lewis, P. J. Hurst, "An Analog Background Calibration Technique for Time-Interleaved Analog-to-Digital Converters," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1912–1919, Dec. 1998. 56, 57, 186

[100] D. Fu, K. C. Dyer, S. H. Lewis, P. J. Hurst, "A Digital Background Calibration Technique for Time-Interleaved Analog-to-Digital Converters," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1904–1911, Dec. 1998. 56, 57

[101] T. Matsuura, T. Nara, T. Komatsu, E. Imaizumi, T. Matsutsuru, R. Horita, H. Katsu, S. Suzumura, K. Sato, "A 240-Mbps, 1-W CMOS EPRML Read-Channel LSI Chip Using an Interleaved Subranging Pipeline A/D Converter," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1840–1850, Nov. 1998. 56, 57

[102] M. K. Mayes, S. W. Chin, L. L. Stoian, "A Low-power 1 MHz, 25 mW 12-Bit Time-Interleaved Analog-to-Digital Converter," *IEEE J. Solid-State Circuits*, vol. 31, pp. 169–178, Feb. 1996. 56

[103] S. Lindfors, T. Rahkonen, M. Waltari, "Digital Correction of Clock Skew in Parallel A/D Converters," *Prestudy report for Correction of Dynamic Errors in Data Converters (EXSITE research program),* Sep. 2000. 57

[104] A. M. Abo, P. R. Gray, "A 1.5-V, 10-bit, 14-MS/s CMOS Pipeline Analog-to-Digital Converter," *IEEE J. Solid-State Circuits*, vol. 34, pp. 599–606, May 1999. 58, 83, 85, 233, 238

[105] B. Razavi, " A 200-MHz 15-mW BiCMOS Sample-and-Hold Amplifier with 3 V Supply," *IEEE J. Solid-State Circuits*, vol. 30, pp. 1326–1332, Dec. 1995. 60

[106] P. Vorenkamp, J. P. M. Verdaasdonk, "Fully Bipolar, 120-Msample/s 10 -b Track-and-Hold Circuit," *IEEE J. Solid-State Circuits*, vol. 27, pp. 988–992, Jul. 1992. 60, 61

[107] B. Prégardier, U. Langmann, W. J. Hillery, "A 1.2-GS/s Silicon Bipolar Track&Hold IC," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1336–1339, Sep. 1996. 61

[108] T. Baumheinrich, B. Prégardier, U. Langmann, "A 1-GSample/s 10-b Full Nyquist Silicon Bipolar Track&Hold IC," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1951–1960, Dec. 1997. 61

[109] C. Fiocchi, U. Gatti, F. Maloberti, "A 10b 250MHz BiCMOS Track and Hold," in *1997 IEEE International Solid-State Circuits Conference, Dig. Tech. Pap*, pp. 144–145, 1997. 61

[110] A. N. Karanicolas, "A 2.7-V 300-MS/s Track-and-Hold Amplifier," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1961–1967, Dec 1997. 61

[111] B. Razavi, "Design of a 100-MHz 10-mW 3-V Sample-and-Hold Amplifier in Digital Bipolar Technology," *IEEE J. Solid-State Circuits*, vol. 30, pp. 724–730, Jul. 1995. 61

[112] K. Hadidi, M. Sasaki, T. Watanabe, D. Muramatsu, T. Matsumoto, "A 103MHz Open-Loop Full CMOS Highly-Linear Sample-and-Hold Amplifier," in *proc European Solid-State Circuits Conference*, pp. 396–399, 1997. 63

[113] B. Razavi, "Design of Sample-and-Hold Amplifiers for High-Speed Low-Voltage A/D Converters," in proc *IEEE 1997 Custom Integrated Circuits Conference*, pp. 59–66, 1997. 63, 64

[114] P. J. Lim, B. A. Wooley, "A High-Speed Sample-and-Hold Technique Using a Miller Hold Capacitance," *IEEE J. Solid-State Circuits*, vol. 26, pp. 643–651, Apr. 1991. 65, 66

[115] "Applications of the CA3080 High-Performance Operational Transconductance Amplifiers," *Harris Semiconductor Application Note 6668*, Nov. 1996. 66

[116] L. Dai, R. Harjani, "CMOS Switched-Opamp Based Sample-and-Hold Circuit," in *Proc. IEEE International Symposium on Circuits and Systems* 1998, vol. I, pp. 476–479. 66, 67

[117] M. Ishikawa, T. Tsukahara, "An 8-bit 50-MHz CMOS Subranging A/D Converter with Pipelined Wide-Band S/H," *IEEE J. Solid-State Circuits*, vol. 24, pp. 1485–1491, Dec. 1989. 67, 68

[118] L. Schillaci, A. Baschirotto, R. Castello, "A 3-V 5.4-mW BiCMOS Track&Hold Circuit with Sampling Frequency up to 150 MHz," *IEEE J. Solid-State Circuits*, vol. 32, pp. 926–932, Jul. 1997. 68

[119] F.-J. Wang, G. C. Temes, "A Fast Offset-Free Sample-and-Hold Circuit," in proc *IEEE 1988 Custom Integrated Circuits Conference*, pp. 5.6.1–5.6.3, 1997. 69

[120] G. Nicollini, P. Confalonieri, D. Senderowicz, "A Fully Differential Sample-and-Hold Circuit for High-Speed Applications," *IEEE J. Solid-State Circuits*, vol. 24, pp. 1461–1465, Oct. 1989. 69, 100, 101

[121] K. Y. Kim, *A 10-bit, 100 MS/s Analog-to-Digital Converter in 1-µm CMOS,* Ph.D. Dissertation, University of California, Los Angeles, 1996. 69, 123

[122] M. de Wit, "Sample and hold circuitry and methods," U.S. Patent 5 170 075, Texas Instruments Inc., Dec. 8, 1992. 73, 83

[123] B. J. Sheu, C. Hu, "Switch-Induced Error Voltage on a Switched Capacitor," *IEEE J. Solid-State Circuits*, vol. sc-19, pp. 519–525, Aug. 1984. 73, 74

[124] W. B. Wilson, H. Z. Massoud, E. J. Swanson, R. T. George, Jr., R. B. Fair, "Measurement and Modelling of Charge Feedthrough in n-Channel MOS Analog Switches," *IEEE J. Solid-State Circuits*, vol. sc-20, pp. 1206–1213, Dec. 1985. 73, 74

[125] B. J. Sheu, J-H. Shieh, M. Patil, "Modeling Charge Injection in MOS Analog Switches," *IEEE Trans. Circuits and Systems*, vol. cas-34, pp. 214–216, Feb. 1987. 73, 76

[126] G. Wegmann, E .A. Vittoz, F. Rahali, "Charge Injection in Analog MOS Switches," *IEEE J. Solid-State Circuits*, vol. sc-22, pp. 1091–1097, Dec. 1987. 73, 74

[127] *BSIM3v3 Manual*, University of California, Berkeley, 1996. 78

[128] D. G. Haigh, B. Singh, "A switching scheme for switched capacitor filters which reduces the effect of parasitic capacitances associated with switch control terminals," in *Proc. IEEE International Symposium on Circuits and Systems*, pp. 586–589, 1983. 78

[129] P. E. Allen, D. R. Holberg, *CMOS Analog Circuit Design*, Holt, Rinehart and Winston, 1987. 80

[130] S. S. Lum, W. C. Rempfer, "Circuits and Methods for Compensating Non-Linear Capacitances to Minimize Harmonic Distortion," U.S. Patent 5 763 924, Linear Technology Corporation, Jun. 9, 1998. 81

[131] T. B. Cho, *Low-Power, Low-Voltage Analog-to-Digital Conversion Techniques using Pipelined Architectures*, Ph.D. Dissertation, University of California, Berkeley, 1995. 82

[132] Y. Nakagome, H. Tanaka, K. Takeuchi, E. Kume, Y. Watanabe, T. Kaga, Y. Kawamoto, F. Murai, R. Izawa, D. Hisamoto, T. Kisu, T. Nishida, E. Takeda,

K. Itoh, "An Experimental 1.5-V 64-Mb DRAM," *IEEE J. Solid-State Circuits*, vol. 26, pp. 465–478, Apr. 1991. 82

[133] C-Y. Wu, W-S. Wey, T-C Yu, "A 1.5 V CMOS Balanced Differential Switched-Capacitor Filter with Internal Clock Boosters," in *Proc. IEEE International Symposium on Circuits and Systems*, pp. 1025–1028, 1995. 83

[134] L. Singer, T. L. Brooks, "Two-Phase Bootstrapped CMOS Switch Drive Technique and Circuit," U.S. Patent 6 118 326, Analog Devices Inc., Sep. 12, 2000. 83

[135] A. M. Abo, P. R. Gray, "A 1.5V, 10-bit, 14MS/s CMOS Pipeline Analog-to-Digital Converter," *1998 Symposium on VLSI Circuits Digest of Technical Papers*, pp. 166–169. 85

[136] M. Dessouky, A. Kaiser, "Input switch configuration for rail-to-rail operation of switched opamp circuits," *Electronics Letters*, vol. 35, no. 1, pp. 8–10, Jan 1999. 86, 152, 181, 220

[137] D. J. Sauer, "Constant Impedance Sampling Switch for an Analog to Digital Converter," U.S. Patent 5 500 612, David Sarnoff Research Center Inc., Mar. 19, 1996. 86

[138] J. L. Gorecki, "Dynamic Input Sampling Switch for CDACS," U.S. Patent 5 084 634, Burr-Brown Corporation, Jan. 28, 1992. 86

[139] J. R. Naylor, M. A. Shill, "Bootstrapped FET Sampling Switch," U.S. Patent 5 172 019, Burr-Brown Corporation, Dec. 15, 1992. 86

[140] T. L. Brooks, D. H. Rebertson, D. F. Kelly, A. D. Muro, S. W. Harston, "A Cascaded Sigma-Delta A/D Converter with 1.25 MHz Signal Bandwidth and 89 dB SNR," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1896-1906, Dec. 1997. 86

[141] L. Singer, T. L. Brooks, "Two-Phase Bootstrapped CMOS Switch Drive Technique and Circuit," U.S. Patent 6 060 937, Analog Devices Inc., May 9, 2000. 86

[142] J. Blake, A. Gribben, C. Price, "Back Gate Switched Sample and Hold Circuit," U.S. Patent 5 422 583, Jun. 6, 1995. 86

[143] J. Steensgaard "Bootstrapped Low-Voltage Analog Switches," in *Proc. IEEE International Symposium on Circuits and Systems*, vol. II, pp. 29–32, 1999. 90

[144] H. Pan, M. Segami, M. Choi, J. Cao, A. Abidi, "A 3.3-V 12-b 50-MS/s A/D Converter in 0.6-μm CMOS with over 80-dB SFDR," *IEEE J. Solid-State Circuits*, vol. 35, pp. 1769-1780, Dec. 2000. 90

[145] M. Dessouky, A. Kaiser, "Very Low-Voltage Digital-Audio $\Delta\Sigma$ Modulator with 88-dB Dynamic Range Using Local Switch Bootstrapping," *IEEE J. Solid-State Circuits*, vol. 36, pp. 349-355, Mar. 2001. 90, 148

[146] H. O. Johansson, C. Svensson, "Time Resolution of NMOS Sampling Switches Used on Low-Swing Signals," *IEEE J. Solid-State Circuits*, vol. 33, pp. 237–245, Feb. 1998. 94

[147] H. C. Yang, D. J. Allstot, "Considerations for Fast Settling Operational Amplifiers," *IEEE Trans. Circuits and Systems,* vol. 37, pp. 326–334, Mar. 1990. 97

[148] A. A. Abidi, "High-Frequency Noise Measurements on FET's with Small Dimensions," *IEEE Trans. Electron Devices*, vol. 33, pp. 1801–1805, Nov. 1986. 99

[149] D. P. Triantis, A. N. Birbas, D. Kondis, "Thermal Noise Modeling for Short-Channel MOSFET's," *IEEE Trans. Electron Devices*, vol. 43, pp. 1950–1955, Nov. 1996. 99

[150] G. Knoblinger, P. Klein, M. Tiebout, "A New Model for Thermal Channel Noise of Deep-Submicron MOSFETS and its Appliction in RF-CMOS Design," *IEEE J. Solid-State Circuits*, vol. 36, pp. 831–837, May 2001. 99

[151] T. C. Choi, R. T. Kaneshiro, W. Brodersen, P. R. Gray, W. B. Jett, M. Wilcox, "High-Frequency CMOS Switched-Capacitor Filters for Communications Application," *IEEE J. Solid-State Circuits*, vol. SC-18, pp. 652–664, Dec. 1983. 101

[152] J. H. Huijsing, F. Tol, "Monolithic Operational Amplifier Design with Improved HF Behavior," *IEEE Journal of Solid-State Circuits*, vol. 11, pp. 323–327, Apr. 1976. 102

[153] W. Sansen, Z. Y. Chang, "Feedforward Compensation Techniques for High-Frequency CMOS Amplifiers," *IEEE Journal of Solid-State Circuits*, vol. 25, pp. 1590–1595, Dec. 1990. 102

[154] F. Op't Eynde, W. Sansen, "A CMOS Wideband Amplifier with 800 MHz Gain-Bandwidth," *IEEE 1991 Custom Integrated Circuits Conference*, pp. 9.1.1–9.1.4, 1991. 102

[155] S. Setty, C. Toumazou, "Feedforward Compensation Techniques in the Design of Low Voltage Opamps and OTAs," *IEEE International Symposium on Circuits and Systems*, vol. I, pp. 464–467, 1998. 102

[156] B. Y. Kamath, R. G. Meyer, P. R. Gray, "Relationship Between Frequency Response and Settling Time of Operational Amplifiers," *IEEE Journal of Solid-State Circuits*, vol. 9, pp. 347–352, Dec. 1974. 102

[157] R. E. Vallee, E. I. El-Masry, "A Very High-Frequency CMOS Complementary Folded Cascode Amplifier," *IEEE Journal of Solid-State Circuits*, vol. 29, pp. 130–133, Feb. 1994. 102

[158] D. A Johns, K. Martin, *Analog Integrated Circuit Design*, John Wiley and Sons, 1997. 102

[159] A. A. Abidi, "On the Operation of Cascode Gain Stages," *IEEE J. Solid-State Circuits*, vol. 23, pp. 1434–1437, Dec. 1988. 102

[160] L. Sumanen, M. Waltari, K. Halonen, "A Low Power A/D Converter for WCDMA Testbed," in *proc NORCHIP'98*, pp. 125–130, Nov. 1998. 102

[161] A. Pärssinen, J. Jussila, J. Ryynänen, L. Sumanen, K. Halonen, "A 2-GHz Wide-Band Direct Conversion Receiver for WCDMA Applications," *IEEE J. Solid-State Circuits*, vol. 34, pp. 1893–1903, Dec. 1999. 102

[162] B. J. Hosticka, "Improvement of the Gain of MOS Amplifiers," *IEEE J. Solid-State Circuits*, vol. 14, pp. 1111–1114, Dec. 1979. 105

[163] E. Säckinger, W. Guggenbühl, "A High-Swing, High-Impedance MOS Cascode Circuit," *IEEE J. Solid-State Circuits*, vol. 25, pp. 289–297, Feb. 1990. 105

[164] U. Gatti, F. Maloberti, G. Torelli, "A Novel CMOS Linear Transcoductance Cell for Continuous-Time Filters," *IEEE International Symposium on Circuits and Systems*, pp. 1173–1176, 1990. 105

[165] S. Zarabadi, M. Ismail "High density integrated circuit with high output impedance," U.S. Patent 5 337 021, Aug. 9, 1994. 105

[166] K. Bult, G. J. G. M. Geelen, "A Fast-Settling CMOS Op Amp for SC Circuits with 90-dB DC Gain," *IEEE J. Solid-State Circuits*, vol. 25, pp. 1379–1384, Dec. 1990. 105

[167] D. Flandre, A. Viviani, J-P. Eggermont, B. Gentinne, P. G. A. Jespers, "Improved Synthesis of Gain-Boosted Regulated-Cascode CMOS Stage Using Symbolic Analysis and gm/ID Methodology," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1006–1012, Jul. 1997. 106

[168] A. M. Marques, V. Peluso, M. S. J. Steyaert, W. Sansen, "A 15-b Resolution 2-MHz Nyquist Rate ΔΣ ADC in a 1-μm CMOS Technology," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1065–1075, Jul. 1998. 106

[169] A. A. Abidi, "An analysis of bootstrapped gain enhancement techniques," *IEEE J. Solid-State Circuits*, vol. 22, pp. 1200–1204, Dec. 1987. 106

[170] C. A. Laber, P. R. Gray, "A Positive-Feedback Transconductance Amplifier with Applications to High-Frequency, High-Q CMOS Switched-Capacitor Filters," *IEEE J. Solid-State Circuits*, vol. 23, pp. 1370–1378, Dec. 1988. 106

[171] B. J. Hosticka, "Dynamic CMOS Amplifiers," *IEEE J. Solid-State Circuits*, vol. 15, pp. 887–894, Oct. 1980. 106

[172] K. Nagaraj, T. R. Viswanathan, K. Singhal, J. Vlach, "Switched-Capacitor Circuits with Reduced Sensitivity to Amplifier Gain," *IEEE Trans. Circuits and Systems*, vol. cas-34, pp. 571–574, May 1987. 106

[173] P. C. Yu, H-S. Lee, "A High-Swing 2-V CMOS Operational Amplifier with Replica-Amp Gain Enhancement," *IEEE J. Solid-State Circuits*, vol. 28, pp. 1265–1272, Dec. 1993. 106

[174] A. R. Feldman, B. E. Boser, P. R. Gray, "A 13-Bit, 1.4-MS/s Sigma-Delta Modulator for RF Baseband Channel Applications," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1462–1469, Dec. 1998. 109

[175] B. K. Ahuja, "An Improved Frequency Compensation Technique for CMOS Operational Amplifiers," *IEEE J. Solid-State Circuits*, vol. 18, pp. 629–633, Dec. 1983. 109

[176] D. B. Ribner, M. A. Copeland, "Design Techniques for Cascoded CMOS Op Amps with Improved PSRR and Common-Mode Input Range," *IEEE J. Solid-State Circuits*, vol. 19, pp. 919–925, Dec. 1984. 109

[177] R. Hogervorst, J. H. Huijsing, *Design of Low-Voltage, Low-Power Operational Amplifier Cells*, Kluwer Academic Publishers, 1996. 110

[178] M. Gustavsson, *CMOS A/D Converters for Telecommunications*, Ph.D. Thesis, Linköping University, 1998. 110

[179] D. Senderowicz, S. F. Dreyer, J. H. Huggins, C. F. Rahim, C. A. Laber, "A Family of Differential NMOS analog Circuits for a PCM Codec Filter Chip," *IEEE J. Solid-State Circuits*, vol. 17, pp. 1014–1023, Dec. 1982. 110, 142

[180] T. H. Lee, A. Hajimiri, "Oscillator Phase Noise: A Tutorial," *IEEE J. Solid-State Circuits*, vol. 35, pp. 326–336, Mar. 2000. 113

[181] S. Lindfors, A. Pärssinen, J. Ryynänen, K. Halonen, "A Novel Technique For Noise Reduction in CMOS Subsamplers," *IEEE International Symposium on Circuits and Systems*, vol. I, pp. 257–260, May 1998. 114

[182] A. Hajimiri, S. Limotyrakis, T. H. Lee, "Jitter and Phase Noise in Ring Oscillators," *IEEE J. Solid-State Circuits*, vol. 34, pp. 790–804, Jun. 1999. 115

[183] T. C. Weigandt, B. Kim, P. R. Gray, "Analysis of Timing Jitter in CMOS Ring Oscillators," *IEEE International Symposium on Circuits and Systems*, pp. 27–30, May 1994. 116

[184] T. C. Choi, R. W. Brodersen, "Considerations for High-Frequency Switched-Capacitor Ladder Filters," *IEEE Trans. Circuits and Systems*, vol. cas-27, pp. 545–552, Jun 1980. 121

[185] D. Senderowicz, G. Nicollini, S. Pernici, A. Nagari, P. Confalonieri, C. Dallavalle, "Low-Voltage Double-Sampled ΔΣ Converters," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1907–1909, Dec. 1997. 121, 126

[186] S. Bazarjani, M. Snelgrove, "A 40 MHz Double-Sampled SC Bandpass ΣΔ Modulator," in *Proc. IEEE International Symposium on Circuits and Systems*, 1997, pp. 73–76. 121, 123

[187] A. Baschirotto, "A 40MHz CMOS Sample&Hold operating at 1.2V," in *proc. 24th European Solid-State Circuits Conference*, pp. 248–251, 1998. 121

[188] Y-C Jenq, "Digital Spectra of Nonuniformly Sampled Signals: Fundamentals and High-Speed Waveform Digitizers," *IEEE Trans. Instrumentation and Measurement*, vol. 37, pp. 245–251, Jun. 1988. 123

[189] J. J. F. Rijns, H. Wallinga, "Spectral Analysis of Double-Sampling Switched-Capacitor Filters," *IEEE Trans. Circuits and Systems*, vol. cas-38, pp. 1269–1279, Nov. 1991. 123

[190] M. Gustavsson, N. N. Tan, "A Global Passive Sampling Technique for High-Speed Switched-Capacitor Time-Interleaved ADCs," *IEEE Trans. Circuits and Systems—II*, vol. 47, pp. 821–831, Sep. 2000. 132

[191] M. Steyaert, J. Crols, S. Gogaert, "Switched-opamp, a technique for realizing full CMOS switched-capacitor filters at very low-voltages," in *proc. 19th European Solid-State Circuit Conference*, pp. 178–181, Sep. 1993. 134

[192] A. Baschirotto, R. Castello, F. Montecchi, "Design strategy for low-voltage SC circuits," *IEE Electronics Letters*, vol. 30, pp. 378–379, Mar. 3, 1994. 134, 135

[193] A. Baschirotto, R. Castello, "A 1-V 1.8-MHz CMOS Switched-Opamp SC Filter with Rail-to-Rail Output Swing," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1979–1986, Dec. 1997. 134, 137, 139, 141, 142, 152

[194] V. Peluso, M. S. J. Steyaert, W. Sansen, "A 1.5-V-100-μW $\Delta\Sigma$ Modulator with 12-b Dynamic Range Using the Switched-Opamp Technique," *IEEE J. Solid-State Circuits*, vol. 32, pp. 943–951, Jul. 1997. 134, 141, 152

[195] V. Peluso, P. Vancorenland, A. M. Marques, M. S. Steyaert, W. Sansen, "A 900-mV Low-Power $\Delta\Sigma$ A/D Converter with 77-dB Dynamic Range," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1887–1897, Dec. 1998. 134, 137, 140, 143, 152

[196] S. Crapanzano, D. A. Johns, "2V fully differential SC integrator in standard CMOS," *IEE Electronics Letters*, vol. 31, no. 23, pp. 1995–1996, Nov. 9, 1995. 137

[197] V. S.-L. Cheung, H. C. Luong, W.-H. Ki, "A 1-V CMOS Switched-Opamp Switched-Capacitor Pseudo-2-Path Filter," *IEEE J. Solid-State Circuits*, vol. 36, pp. 14–22, Jan. 2001. 139

[198] A. Baschirotto, R. Castello, "A 1-V 1.8-MHz CMOS Switched-Opamp SC Filter with Rail-to-Rail Output Swing," *1997 IEEE International Solid-State Circuits Conference, Dig. Tech. Pap.*, pp. 58–59, Feb. 1997. 141

[199] V. Peluso, P. Vancorenland, M. Steyaert, W. Sansen, "900mV differential class AB OTA for switched opamp applications," *IEE Electronics Letters*, vol. 33, no. 17, pp. 1455–1456, Aug. 14, 1997. 143

[200] R. J. Baker, H. W. Li, D. E. Boyce, *CMOS Circuit Design, Layout and Simulation*, IEEE Press, 1998. 146

[201] J. Sauerbrey, R. Thewes, "Ultra Low Voltage Switched Opamp ΣΔ Modulator for Portable Applications," *IEEE 2001 Custom Integrated Circuits Conference*, pp. 35–38, 2001. 148

[202] A. Baschirotto, R. Castello, "A 1V CMOS fully-differential swtched-opamp bandpass ΣΔ modulator," in *Proceedings of the 23th European Solid-State Circuits Conference*, pp. 152–155, 1997. 152

[203] S. Karthikeyan, S. Mortezapour, A. Tammineedi, E. K. F. Lee, "Low-Voltage Analog Circuit Design Based on Biased Inverting Opamp Configuration," *IEEE Trans. Circuits and Systems–II*, vol. 47, pp. 176–184, Mar. 2000. 153

[204] S. Mortezapour, E. K. F. Lee, "A 1-V, 8-Bit Succesive Approximation ADC in Standard CMOS Process," *IEEE J. Solid-State Circuits*, vol. 35, pp. 642–646, Apr. 2000. 153

[205] A. Baschirotto, R. Castello, G. P. Montagna, "Active series switch for switched-opamp circuits," *IEE Electronics Letters*, vol. 34, no. 14, pp. 1365–1366, Jul. 9, 1998. 154, 155

[206] S. Karthikeyan, A. Tamminneedi, C. Boecker, E. K. F. Lee, "Design of Low-Voltage Front-End Interface for Switched-Opamp Circuis," *IEEE Trans. Circuits and Systems–II*, vol. 48, pp. 722–726, Jul. 2001. 154

[207] M. Waltari, S. Lindfors, K. Halonen, "A Highly Linear V-I Converter," in *proc European Conference on Circuit Theory and Design*, vol. I, pp. 98–101, Aug 1999. 154

[208] E. Bidari, M. Keskin, F. Maloberti, U. Moon, J. Steensgaard, G. C. Temes, "Low -voltage switched-capacitor circuits," *IEEE International Symposium on Circuits and Systems*, vol. V, pp. 445–448, May 2000. 159, 160

[209] L. Wang, S. H. K. Embabi, E. Sanchez-Sinencio, "1.5V 5.0MHz Switched Capacitor Circuits in 1.2μm CMOS Without Voltage Bootstrapper," in *proc IEEE 2001 Custom Integrated Circuits Conference*, pp. 2.4.1–2.4.4, 2001. 160

[210] M. Keskin, U.-K. Moon, G. C. Temes, "A 1-V, 10-MHz Clock-Rate, 13-Bit CMOS ΔΣ Modulator Using Unity-Gain-Reset Opamps," in *Proceedings of the 27th European Solid-State Circuits Conference*, pp. 532–535, 2001. 160

[211] J. Ramírez-Angulo, R. G. Carvajal, J. Tombs, A. Torralba, "Simple technique for opamp continuous 1V supply operation," *IEE Electronics Letters*, vol. 35, no. 4, pp. xxx–yyy, Feb. 18, 1999. 161

[212] D. R. Welland, "Transconductance amplifiers and exponential variable gain amplifiers using the same," U.S. Patent 5 451 901, Sep. 19, 1995. 162

[213] R. Ahola, S. Lindfors, J. Routama, K. Halonen, "A Novel Phase Detector with No Dead Zone and a Chargepump with Very Wide Output Voltage Range," in *Proceedings of the 24th European Solid-State Circuits Conference*, pp. 352–355, 1998. 162

[214] F. You, S. H. K. Embabi, J. F. Duque-Carillo, E. Sanchez-Sinencio, "An Improved Tail Current Source for Low Voltage Applications," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1173–1179, Aug. 1997. 162

[215] K. Gulati, H.-S. Lee, "A High-Swing Telescopic Operational Amplifier," *IEEE J. Solid-State Circuits*, vol. 33, pp. 2010–2019, Dec. 1998. 162

[216] H. Neuteboom, B. M. J. Kup, M. Jansens, "A DSP-Based Hearing Instrument IC," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1790–1806, Nov. 1997. 163

[217] H. Banba, H. Shiga, S. Umezawa, T. Miyaba, T. Tanzawa, S. Atsumi, K. Sakui, "A CMOS Bandgap Reference Circuit with Sub-1-V Operation," *IEEE J. Solid-State Circuits*, vol. 34, pp. 670–674, May 1999. 163

[218] Y. Jiang, E. K. F. Lee, "Design of Low-Voltage Bandgap Reference Using Transimpedance Amplifier," *IEEE Trans. Circuits and Systems–II*, vol. 47, pp. 552–555, Jun. 2000. 163, 164

[219] P. Malcovati, F. Maloberti, C. Fiocchi, M. Pruzzi, "Curvature-Compensated BiCMOS Bandgap with 1-V Supply Voltage," *IEEE J. Solid-State Circuits*, vol. 36, pp. 1076–1081, Jul. 2001. 163, 164

[220] *IEEE-STD-1241-2000 – Standard for Terminology and Test Methods for Analog-to-Digital Converters* 167, 168

[221] *IEEE-STD-1057-1994 – Standard for Digitizing Waveform Recorders* 168

[222] J. Blair, "Histogram Measurement of ADC Nonlinearities Using Sine Waves," *IEEE Trans. Instrum. Meas.*, vol. 43, pp. 373–383, Jun. 1994. 168

[223] B. E. Peetz, A. S. Muto, J. M. Neil, "Measuring Waveform Recorder Performance," *Hewlett-Packard Journal*, pp. 21–29, Nov. 1982. 168

[224] B. E. Peetz, "Dynamic Testing of Waveform Recorders," *IEEE Trans. Instrum. Meas.*, vol. IM-32, pp. 12–17, Mar. 1983. 168

[225] J. Doernberg, H.-S. Lee, D. A. Hodges, "Full-Speed Testing of A/D Converters," *IEEE J. Solid-State Circuits*, vol. SC-19, pp. 820–827, Dec. 1984. 168

[226] M. Vanden Bossche, J. Schoukens, J. Renneboog, "Dynamic Testing and Diagnostics of A/D Converters," *IEEE Trans. Circuits and Systems*, vol. CAS-33, pp. 775–785, Aug. 1986. 168

[227] J. Yuan, C. Svensson, "New Single-Clock CMOS Latches and Flipflops with Improved Speed and Power Savings," *IEEE J. Solid-State Circuits*, vol. 32, pp. 62–69, Jan. 1997. 171

[228] L. Sumanen, M. Waltari, V. Hakkarainen, K. Halonen, "CMOS Dynamic Comparators for Pipeline A/D Converters," accepted in *ISCAS'02*, May 2002. 184

[229] R. Jewett, K. Poulton, K.-C. Hsieh, J. Doernberg, "A 12b 128MSample/s ADC with 0.05LSB DNL," *ISSCC Dig. Tech. Papers*, pp. 138–139, Feb. 1997. 187

[230] A. R. Bugeja, B-S. Song, "A Self-Trimming 14-b 100-MS/s CMOS DAC," *IEEE J. Solid-State Circuits*, vol. 35., pp. 1841–1852, Dec. 2000. 213

[231] G. A. M. Van der Plas, J. Vandenbussche, W. Sansen, M. S. J. Steyaert, G. G. E. Gielen, "A 14-bit Intrinsic Accuracy Q^2 *Random Walk* CMOS DAC," *IEEE J. Solid-State Circuits*, vol. 34., pp. 1708–1718, Dec. 1999. 213

[232] K.-C. Hsieh, T. A. Knotts, G. L. Baldwin, T. Hornak, "A 12-bit 1-Gword/s GaAs Digital-to-Analog Converter System," *IEEE J. Solid-State Circuits*, vol. SC-22., pp. 1048–1055, Dec. 1987. 213

[233] A. R. Bugeja, B-S. Song, P. L. Rakers, S. F. Gillig, "A 14-b, 100-MS/s CMOS DAC Designed for Spectral Performance," *IEEE J. Solid-State Circuits*, vol. 34., pp. 1719–1732, Dec. 1999. 213

[234] D. G. Nairn, A. Biman, "A Comparative Analysis of Switched-Current Circuits," *IEEE Trans. Circuits Systs. II*, vol. 43, pp. 733–743, Nov. 1996. 218

[235] D. W. J. Groeneveld, H. J. Schouwenaars, H. A. H. Termeer, C. A. A. Bastiaansen, "A Self-Calibration Technique for Monolithic High-Resolution D/A

Converters," *IEEE J. Solid-State Circuits*, vol. 24., pp. 1517–1522, Dec. 1989. 218

[236] G. Wegmann, E. A. Vittoz, "Analysis and Improvements of Accurate Dynamic Current Mirrors," *IEEE J. Solid-State Circuits*, vol. 25., pp. 699–706, Jun. 1990. 218

[237] J. B. Hughes, K. W. Moulding, "S^{2}I: A Switched-Current Technique for High Performance," *IEE Electronics Letters*, vol. 29, no. 16, pp. 1400–1401, Aug 1993. 219

[238] D. G. Nairn, "A High-Linearity Sampling Technique for Switched-Current Circuits," *IEEE Trans. Circuits Systs. II*, vol. 43, pp. 49–52, Jan. 1996. 219

[239] S. Hadri, B. Leung, "Impedance Boosting Techniques Based on BiCMOS Technology," *IEEE J. Solid-State Circuits*, vol. 28., pp. 157–161, Feb. 1993. 220

[240] C. C. Chen, C. Y. Wu, "Design Techniques for 1.5-V Low-Power CMOS Current-Mode Cyclic Analog-to-Digital Converters," *IEEE Trans. Circuits Syst. II*, vol. 45, pp. 28–40, Jan. 1998. 224

[241] V. Koifman, Y. Afek, J. Shor, "Circuit Methods for the Integration of Low Voltage (1.1–1.8V) Analog Functions on System-on-a-Chip IC's in a Single-Poly CMOS Processes," in proc. *ISLPED'99,* pp. 60–63, Aug. 1999. 224

[242] L. Sumanen, M. Waltari, K. Halonen, "A Pipeline A/D Converter for WCDMA Testbed," *Analog Integrated Circuits and Signal Processing*, vol. 22, pp. 41–50, Jan. 2000. 265

Appendix A

Derivation of OTA GBW Requirement

The small signal model for an SC amplifier in hold mode is shown in Figure A.1, where the OTA is represented with a single pole model consisting of g_m, g_o, and C_L. The capacitance C_L is the sum of the OTA output capacitance and the external load capacitance. The capacitance at the OTA input is represented with C_{in} and the output conductance with g_o.

This circuit is used to find out the settling time constant using a pulsed current source i_i as the excitation signal. The small signal analysis yields the following transfer function:

$$\frac{v_o}{i_i} = -\frac{g_m - sC_F}{s\left[(g_mC_F + g_oC_{i,tot} + g_oC_F) + s(C_{i,tot}C_L + C_LC_F + C_FC_{i,tot})\right]}, \qquad \text{(A.1)}$$
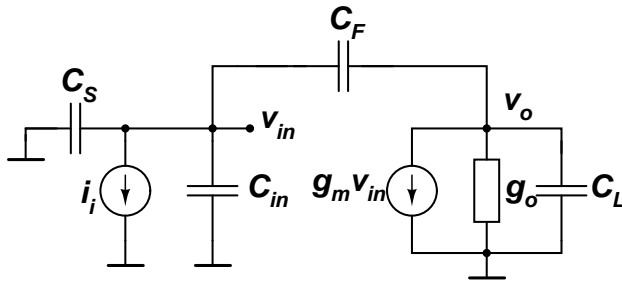
Figure A.1 Small signal model for SC amplifier in hold mode.

where $C_{i,tot} = C_S + C_{in}$. The output voltage for a current impulse with total integrated charge Q can be written (in partial fraction form) as

$$v_o = \frac{g_m Q}{s(g_m C_F + g_o C_{i,tot} + g_o C_F)} - \frac{\frac{C_F}{C_{i,tot} C_L + C_L C_F + C_F C_{i,tot}} + \frac{g_m}{g_m C_F + g_o C_{i,tot} + g_o C_F}}{\frac{g_m C_F + g_o C_{i,tot} + g_o C_F}{C_{i,tot} C_L + C_L C_F + C_F C_{i,tot}} + s} Q. \quad (A.2)$$

The corresponding time domain expression is obtained with the inverse Laplace transform, giving

$$
\begin{aligned}
v_0 &= \frac{Q}{C_F + \frac{g_o}{g_m} C_{i,tot} + \frac{g_o}{g_m} C_F} \cdot \left(1 - e^{-\frac{g_m C_F + g_o C_{i,tot} + g_o C_F}{C_{i,tot} C_L + C_L C_F + C_F C_{i,tot}} t}\right) \\
&\quad - \frac{C_F Q}{C_{i,tot} C_L + C_L C_F + C_F C_{i,tot}} \cdot e^{-\frac{g_m C_F + g_o C_{i,tot} + g_o C_F}{C_{i,tot} C_L + C_L C_F + C_F C_{i,tot}} t},
\end{aligned}
\quad (A.3)
$$

from where the settling time constant can be identified as

$$\tau = \frac{C_{i,tot} C_L + C_L C_F + C_F C_{i,tot}}{g_m C_F + g_o C_{i,tot} + g_o C_F}. \quad (A.4)$$

Since g_m is large compared to g_o the time constant can be approximated with

$$\tau \approx \frac{C_{i,tot} C_L + C_L C_F + C_F C_{i,tot}}{g_m C_F}. \quad (A.5)$$

A well-known relation states that the gain-bandwidth product of a single-pole OTA is

$$GBW = \frac{g_m}{2\pi C_L}. \quad (A.6)$$

Substituting this to (A.5) yields

$$\tau = \frac{1}{2\pi GBW} \cdot \left(1 + \frac{C_{i,tot}(C_L + C_F)}{C_F C_L}\right). \quad (A.7)$$

When the last term resulting from the feed-forward path is ignored in (A.3), the output settles to N-bit accuracy in the time period T if

$$e^{-\frac{T}{\tau}} < 2^{-N}, \quad (A.8)$$

which leads to the following requirement for the GBW:

$$GBW > \frac{N \ln 2 \cdot \left(1 + \frac{C_{i,tot}(C_L + C_F)}{C_F C_L}\right)}{2\pi T}. \quad (A.9)$$

The *GBW* can also be expressed in terms of the feedback factor $f = C_F/(C_F + C_{i,tot})$ and the effective load capacitance $C_{L,tot} = C_L + C_F C_{i,tot}/(C_F + C_{i,tot})$ [242], resulting in

$$GBW > \frac{N \ln 2 C_{L,tot}}{2\pi T f C_L}. \tag{A.10}$$

Appendix B

Optimum Input Capacitance

In this appendix an optimum value, which minimizes the settling time of an SC amplifier, will be derived for OTA input transistor gate capacitance. In strong inversion the transconductance of a MOS transistor is given by

$$g_m = \sqrt{\frac{I_D \mu C_{ox} W}{L}}.$$ (B.1)

Using the gate capacitance $C_G = C_{ox}WL$, it can be rewritten as

$$g_m = \sqrt{\frac{I_D \mu C_G}{L^2}}.$$ (B.2)

The capacitance C_{in} in Appendix A is now C_G. Again, from Appendix A the settling time constant (A.7) becomes

$$\tau = \frac{C_L L}{\sqrt{I_D \mu C_G}} \cdot \left(1 + \frac{(C_S + C_G)(C_L + C_F)}{C_F C_L}\right).$$ (B.3)

Finding the minimum time constant yields the optimum gate capacitance, which is given by

$$C_{G,opt} = \frac{C_F C_L}{C_F + C_L} + C_S.$$ (B.4)

When the same analysis is repeated, assuming that the transistor is biased in the weak inversion, where the transconductance is given by

$$g_m = \frac{nkT}{q} I_D,$$ (B.5)

the settling time constant becomes

$$\tau = \frac{qC_L}{nkTI_D} \cdot \left(1 + \frac{(C_S + C_G)(C_L + C_F)}{C_F C_L}\right). \tag{B.6}$$

Now there is no optimum gate capacitance, but the settling time increases with C_G (i.e. with W when L is fixed).

In velocity saturation the transconductance is given by

$$g_m = v_{sat} C_{ox} W = \frac{v_{sat} C_G}{L}, \tag{B.7}$$

which results in the following settling time constant:

$$\tau = \frac{LC_L}{v_{sat} C_G} \cdot \left(1 + \frac{(C_S + C_G)(C_L + C_F)}{C_F C_L}\right). \tag{B.8}$$

Again, there is no optimum gate capacitance. As opposed to the weak inversion case, the settling time now decreases with increasing C_G (i.e with increasing W when L is fixed).

Appendix C

Saturation Voltage

This appendix will study how the opamp output transistor saturation voltage V_{dsat} can be scaled with respect to V_{DD}. It is assumed that $V_{dsat} = k_6 V_{DD}^m$, and thus the task is to find expressions for the constants m and k_6.

Another assumption is that the transistor current is determined by the slew-rate requirement, and thus $I_D = I_{SR}$. The opamp employs the Miller topology, where the second pole is determined by the output transistor and is given by

$$p_2 = -\frac{g_m C_C}{C_L(C_C + C_1)}, \tag{C.1}$$

where C_C is the compensation capacitance, C_L the load capacitance, and C_1 the first stage output capacitance (including the second stage input capacitance). All the capacitances are assumed to be linearly proportional to the sampling capacitor C and thus $p_2 \propto -g_m/C$. To prevent the opamp from losing speed, the magnitude of p_2 has to be constant or increasing as V_{DD} is scaled down. The pole is given by

$$\frac{g_m}{C} = \frac{2I_D}{(V_{GS} - V_T)C} = \frac{2I_{SR}}{V_{dsat}C}. \tag{C.2}$$

Substituting I_{SR} from (2.8) yields

$$\frac{g_m}{C} \propto \frac{(V_{DD} - V_{margin})}{V_{dsat}} = \frac{V_{DD}}{k_6 V_{DD}^m} - \frac{V_{margin}}{V_{dsat}}. \tag{C.3}$$

Since V_{margin} is proportional to V_{dsat}, which makes the last term constant, the pole frequency will not decrease if $m \geq 1$.

Another important thing is the second stage input capacitance C_1. It is not allowed

to grow faster than the compensation capacitance, i.e. not faster than C. It is assumed that C_1 is dominated by the transistor gate capacitance. From the transistor current equation we can get

$$C_G = C_{ox}WL \propto \frac{I_{SR}L^2}{V_{dsat}^2} \propto \frac{(V_{DD} - V_{margin})CL^2}{k_6^2 V_{DD}^{2m}}. \tag{C.4}$$

For a given technology (fixed L) m must be 0.5 or less, which conflicts with the earlier requirement. If L scales linearly with the supply voltage, then $m \leq 1.5$, which, together with the earlier requirement, results in $1 \leq m \leq 1.5$.

So, it was wrong to assume that in the case of a fixed technology the output transistor size and current are always constrained by the slew rate. Writing the expression for the second pole again yields

$$\frac{g_m}{C} = \frac{2\mu C_{ox}W(V_{GS} - V_T)}{LC} = \frac{2\mu C_G V_{dsat}}{L^2 C}. \tag{C.5}$$

If C_G is now allowed to grow as fast as possible by making it proportional to C, then V_{dsat} will be constant and the current will need to be increased faster than required by the slew rate.

The above analysis is made for a Miller opamp, but the results are applicable to other opamp topologies as well. For example, in the folded cascode opamp the non-dominant pole is determined by the g_m of the cascode transistor and parasitic capacitances that are proportional to the gate capacitance. Thus, an analysis of it would yield similar results.