CMOS A/D Converters
Using MOSFET-Only R-2R Ladders

Diss. ETH No. 13921

CMOS A/D Converters Using MOSFET-Only R-2R Ladders

A dissertation submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of

Doctor of Technical Sciences

presented by

CLEMENS MARIA HAMMERSCHMIED
Dipl. El.-Ing. ETH
born October 1, 1970
citizen of Austria

accepted on the recommendation of

Prof. Dr. Q. Huang, examiner
Prof. Dr. H.-A. Loeliger, co-examiner

2000

Acknowledgements

Verzeih, ich kann nicht hohe Worte machen,
Und wenn mich auch der ganze Kreis verhöhnt;
Mein Pathos brächte dich gewiß zum Lachen,
Hättst du dir nicht das Lachen abgewöhnt.

MEPHISTOPHELES, Prolog im Himmel

When finally a thesis is finished, many individuals other than the author have contributed to the work and also to its success. Although the space allocated to acknowledge the help and suggestions from supervisors and colleagues is only about one percent of the entire thesis, the impact on the work has been much larger.

Prof. Dr. Qiuting Huang not only supervised my thesis with great engagement, but also taught me analog integrated circuit design right from the beginning. One thing I learned during all our discussions besides technical aspects was to look at problems from different points of view and to put them in the right context. Moreover, due to the variety of activities in our group I had the chance to gain insight into several other highly fascinating topics of analog circuit design.

I am also indebted to Prof. Dr. Hans-Andrea Loeliger for co-examining my thesis. His enthusiasm and interest made the discussion of the concepts and ideas a real pleasure. Moreover, he made available his time in a way far beyond what could have been reasonably expected.

I am grateful to Norbert Felber who got me into contact with integrated circuits in the first place and who was always willing to discuss issues on digital circuits, signal processing and measurements.

I would also like to thank my colleagues Christoph Kuratli and Christian Menolfi who shared the office with me for most of the time. The many discussions on various topics of analog design improved the work presented here substantially. Moreover, their friendship made the time at the institute much more pleasant.

Integrated circuits design is closely linked to computer software, often more so than preferable. Christoph Wicki and Andreas Karrer always maintained a professional perfectly working computer environment at the institute. Special thanks for providing *vi* on all machines in a world of more and more fancy graphical user-interface based tools.

It is due to Hanspeter Mathys and Hansjörg Gisler that life in the lab was easy. Whether it was a special mechanical setup, the ordering of hard-to-find electronic components or taking pictures of my integrated circuits, I always found a helping hand.

I would also like to thank my parents who gave me the opportunity to study at the ETH in Zurich and thus set ground for this thesis.

Esther finally deserves the biggest thanks for her love and support. She improved the quality of this thesis to a large extent by carefully proofreading the manuscript. On top of this, she was always there when I needed comfort, and giving someone a smile is often more important than technical advice.

All epigraphs are taken from *Johann Wolfgang von Goethe*: FAUST. EINE TRAGÖDIE. The complete citations as well as the translation of the german verses can be found at the end of the thesis.

Zürich, November 2000 Clemens Hammerschmied

Contents

Abstract

Der Teufel hat hier weiter nichts zu sagen.

MEPHISTOPHELES, Der Tragödie zweiter Teil

This dissertation describes the analysis and design of an R-2R ladder structure using MOS transistors instead of resistors and the employment of this ladder in two A/D converters with different architectures.

An important design issue of A/D converters is the achievement of high accuracy. Special post-processing techniques such as laser trimming can be used to improve the accuracy, but this additional manufacturing step is expensive. Another possibility is to include on-chip trimming or calibration circuits. The disadvantage, however, is the additional area and power consumption needed for these circuits.

To avoid calibration or trimming, the accuracy must be achieved by relying on the matching capabilities of active or passive elements. Often capacitors or resistors are used for matching purposes, but these devices are not always available in modern digital deep submicron processes.

An alternative is to use MOS transistors as matching elements. A circuit that is well suited for this purpose is an R-2R-type ladder structure that is based on MOS transistors instead of resistors. The principle of operation is analyzed in detail. Special attention is paid to second-order effects that affect the accuracy of the binary current weighting. Design guidelines are given to achieve high linearity. Measurements in a 1 μm digital CMOS technology show a linearity of more than 11 bit.

To prove the dynamic performance of this structure, the MOSFET-only ladder is used as a D/A converter in a successive-approximation A/D converter. In this architecture, the accuracy is almost entirely determined by the D/A converter and thus by the ladder. The converter is implemented in a digital 1 µm CMOS technology without precise capacitors or resistors. The resolution is 10 bit at a maximum conversion rate of 200 ksample/s. Measurements show an excellent total harmonic distortion (THD) of −79 dB which is achieved without calibration or trimming.

A second A/D converter implementation is aimed at higher conversion speed. A pipelined architecture is chosen that is well suited for converters with resolutions of more than 10 bit and high sampling rates. As with the successive-approximation A/D converter, the D/A converter is also the most important building block of the pipeline stage in terms of linearity.

The four stage pipelined converter is implemented in a digital 0.25 µm CMOS process. When a deep submicron process with a supply voltage of 2.5 V is used for analog design problems arise due to the low voltage headroom. These difficulties are discussed and solutions are presented.

The resolution of the pipelined A/D converter is 13 bit, the maximum conversion rate equals 5 Msample/s. The measured total harmonic distortion is −70 dB at 5 Msample/s. The spurious-free dynamic range (SFDR) is 76 dB. The differential non-linearity (DNL), which also has been measured at the maximum sampling rate, is 0.5 LSB. The moderate power consumption is 118 mW.

Both A/D converters show that the MOSFET-only ladder is an accurate and versatile building block. Due to the exclusive use of MOS transistors it can be implemented in every standard digital CMOS process and is well suited for modern deep submicron technologies.

Zusammenfassung

Der Teufel hat hier weiter nichts zu sagen.

MEPHISTOPHELES, Der Tragödie zweiter Teil

Diese Dissertation beschreibt die Analyse und das Design einer R-2R Leiterstruktur, die aus MOS-Transistoren statt aus Widerständen besteht, sowie deren Einsatz in zwei Analog-Digital-(A/D)-Wandlern mit unterschiedlichen Architekturen.

Ein wichtiges Design-Ziel bei A/D-Wandlern ist das Erreichen hoher Genauigkeit. Durch spezielle Verfahren wie Lasertrimmung nach der Schaltungsherstellung kann die Genauigkeit erhöht werden, jedoch sind solche Maßnahmen teuer. Eine andere Möglichkeit besteht darin, Kalibrations-Schaltungen auf dem Chip zu integrieren, aber dadurch werden sowohl der Leistungsverbrauch als auch die Chipfläche vergrößert.

Um eine Kalibration zu vermeiden, muß die Linearität durch die relative Genauigkeit von aktiven oder passiven Elementen erreicht werden. Meist werden Widerstände oder Kapazitäten für diesen Zweck verwendet, jedoch sind diese Elemente bei modernen Prozessen mit Kanallängen im Sub-Mikrometerbereich oft nicht verfügbar.

Als Alternative können MOS-Transistoren als genauigkeitsbestimmende Elemente verwendet werden. Eine dafür geeignete Schaltung ist die aus MOS-Transistoren bestehende R-2R Leiter. Eine detaillierte Analyse erläutert das zugrundeliegende Prinzip. Besonders wichtig sind Effekte zweiter Ordnung, die die Genauigkeit der binären Stromteilung

der Leiter beeinflussen. Es werden Design-Richtlinien gegeben, um hohe Genauigkeit zu erreichen. Messungen, die mit Leitern in einem CMOS-Prozeß mit einer Kanallänge von 1 µm durchgeführt wurden, zeigen eine Genauigkeit von mehr als 11 bit.

Um das dynamische Verhalten zu prüfen, wurde die Leiter als Digital-Analog-(D/A)-Wandler in einem A/D-Wandler nach dem Prinzip der sukzessiven Approximation verwendet. Bei dieser Architektur wird die Genauigkeit fast ausschließlich durch den D/A-Wandler und damit durch die Leiter bestimmt. Für diesen Wandler wurde ein 1 µm-Prozeß verwendet, der weder über genaue Kapazitäten noch über genaue Widerstände verfügt. Die Auflösung beträgt 10 bit, die Abtastrate 200 kHz. Messungen zeigen eine ausgezeichnete harmonische Verzerrung von −79 dB, die ohne Kalibration erreicht wurde.

Die Zielsetzung des zweiten A/D-Wandlers ist höhere Geschwindigkeit. Es wurde eine Pipeline-Architektur gewählt, die für hohe Abtastraten bei Auflösungen von über 10 bit Vorzüge aufweist. Wie bei dem ersten Wandler ist auch hier der D/A-Wandler das genauigkeitsbestimmende Element.

Der vierstufige A/D-Wandler wurde in einem Prozeß mit 0.25 µm Kanallänge integriert. Die damit verbundene niedrige Versorgungsspannung von 2.5 V wirft Probleme beim Design von Analogschaltungen auf. Diese Probleme und ihre Lösungen werden diskutiert.

Die Auflösung des A/D-Wandlers beträgt 13 bit, die maximale Abtastrate ist 5 MHz. Die harmonischen Verzerrungen betragen −70 dB bei maximaler Geschwindigkeit. Der störungsfreie Dynamikbereich wurde mit 76 dB gemessen. Die differentielle Nichtlinearität, ebenfalls bei Maximalgeschwindigkeit gemessen, beträgt 0.5 LSB. Der Leistungsverbrauch ist mit 118 mW relativ gering.

Beide A/D-Wandler zeigen, daß die aus MOS-Transistoren bestehende Leiter hohe Genauigkeit erreicht. Da sie nur aus Transistoren besteht, kann sie ohne Einschränkung in jeder digitalen Technologie verwendet werden, vor allem auch in modernen Prozessen mit Kanallängen im Sub-Mikrometerbereich.

1

Introduction

Bedenke wohl die erste Zeile,
Daß deine Feder sich nicht übereile!

FAUST, Der Tragödie erster Teil

Electronic products play a dominant role in today's life. We cannot imagine anymore to manage without the help of electronic tools such as computers, television and hi-fi systems or the possibilities of modern telecommunication. Most of the electronic equipment today is digital, not only computers and networks, but also many things of day-to-day life such as elevators, dishwashers, or the motor management of a car. The world, however, stays analog.

Analog-to-digital (A/D) converters provide the interface between the analog world and digital signal processing systems and are therefore used extensively today. Most quantities that are used as input signals for electronic devices, such as the temperature for an air conditioning unit or the voice of a person for a mobile phone, are analog in nature. To be processed by digital circuitry, however, these signals have to be converted to the digital domain by means of an A/D converter.

This thesis discusses the development and design of A/D converters using MOS transistors as elements defining the accuracy. Therefore, this technique can be used in any standard digital CMOS (Complementary Metal Oxide Semiconductor) technology and is thus universally applicable. In addition, it can be used in modern digital deep submicron processes that are more and more used for analog circuits also to enable the integration of a complete system on a chip.

Before going in medias res, a short introduction to typical applications of A/D converters, their history and design challenges will be given.

1.1 Digital Signal Processing Systems

A typical digital system with analog input and output quantities is depicted in figure 1.1. The bandwidth of the analog input signal is limited to the Nyquist frequency by means of an anti-alias filter. This band-limited signal is then fed to the sample-and-hold stage that performs the first step towards a digital signal. The input signal that has been continuous in time is converted to a discrete-time signal. The amplitude, however, remains continuous after the sample-and-hold stage. The quantization of the amplitude is performed by the A/D converter. While during the conversion of a continuous-time signal to a discrete-time signal no information is lost if the signal is band-limited to the Nyquist frequency, the amplitude quantization that maps a continuous signal to a finite number of discrete values inevitably causes a decrease of information such that a complete reconstruction of the signal is not possible any more. The A/D converter thus limits the accuracy and the dynamic range of the entire system. The design of the converter must therefore be carried out with special care.

The output of the A/D converter consists of digital data that represents the analog input signal apart from the limitations mentioned above. A digital system processes this data. This system can be as simple as a

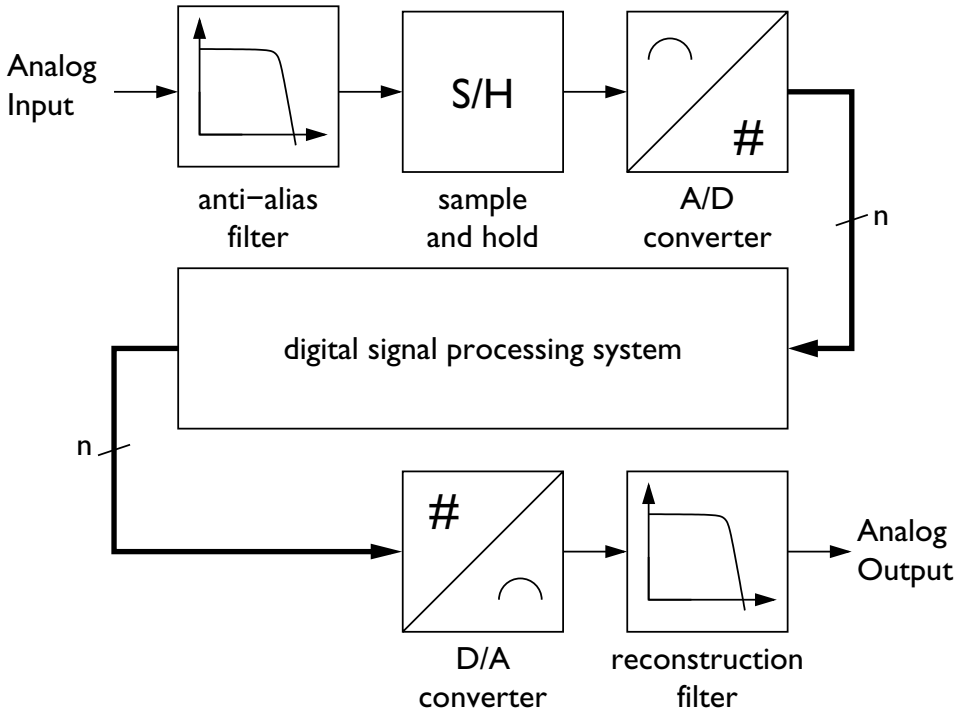

Figure 1.1: Typical block diagram of a digital signal processing system with analog inputs and outputs

single digital filter, but can also be as complex as, e. g., a digital telephone system. In either case the resulting output is also digital data that has to be converted back to the analog domain. This step is performed by the digital-to-analog (D/A) converter, whose output is an analog, but still discrete-time signal. A reconstruction filter finally creates the continuous-time output signal.

Although the entire system seems to be very complex, it has a number of advantages over purely analog signal processing. The most obvious advantage is the flexibility of the system. As mentioned above, the digital signal processing block can perform any function, from very simple to very complex. Many tasks that are complicated to achieve in the analog domain, such as storing large amounts of data, are comparatively

easy in the digital domain. Moreover, the entire system can be made programmable or even adaptive to the current situation.

In an analog signal processing system, the addition of functionality often degrades the signal. Any filtering stage, e. g., that is added to the signal path also adds noise to the signal and thus reduces the dynamic range. In a digital system, however, the losses are small once the analog signal is converted to the digital domain, no matter how complex the operations are that are performed with the digital data.

Some disadvantages also exist, however. Although the processing power of today's microprocessors and digital signal processors (DSPs) increases almost every hour, analog signal processing is still faster in many applications. This is due to the high number of mathematical operations that have to be performed in the digital domain, e. g. for filtering applications. Moreover, the power consumption of the entire signal processing chain including the A/D and D/A converters is often much higher than the power consumed by an analog equivalent. Therefore, analog solutions often are mandatory in low-power applications.

Nevertheless, the advantages of the digital approach are more important in most applications. Systems similar to the one depicted in figure 1.1 are widely used today, which makes the A/D converter an important building block in many applications.

1.2 A/D Converter Development

The first interest on data conversion was shown in the late thirties of this century in the context of pulse code modulation techniques used for telephone communication. But only the appearance of digital computers and the need for data processing systems for avionics and missile programs in the beginning of the fifties led to a fast development of A/D converters [1].

Many of the A/D converter architectures still used today were invented in the fifties and sixties. U.S. patents were issued for some of the

best known A/D converter types during this time. Patents for the flash converter architecture, e. g., were granted in 1959 and 1971 [2, 3], and a patent describing the successive-approximation type converter can be found in 1961 [4].

For a long time, the only A/D converters were of a discrete type, before the process technology matured enough to make the full integration of A/D converters possible. Daniel Sheingold of Analog Devices writes in 1986: "A quarter century ago, A/D converters capable of 0.05% performance and 50,000-sample-per-second conversion rates cost about $8000, consumed about 500 watts, and occupied about one-quarter of a cubic meter. Today, the completely self-contained monolithic Analog Devices AD573 requires less than 20 microseconds for a 10-bit 0.05% conversion, is available in quantity at less than 0.2% of the price, and is packaged in a 20-pin plastic DIP. And it is designed for easy interfacing with the modern microprocessor." [5]

New process technologies enable the development of new types of A/D converters. The pipelined converter, e. g., which is one of the most popular architectures for high-speed converters today, was first integrated in 1987 when the CMOS processes were mature enough to allow the integration of high-speed sample-and-hold circuits [6].

Although the basic A/D converter architectures have been known for some forty years now, the development of A/D converters always was and still is closely linked to new technologies, especially in telecommunications. In the late seventies, the introduction of digital video systems led to a whole new generation of high-speed A/D converters with a resolution of eight to ten bit and a sampling rate exceeding 10 MHz.

In the last couple of years the introduction of new wireline communication technologies such as ADSL (Asymmetric Digital Subscriber Line) or VDSL (Very High Speed Digital Subscriber Line), which provide high-speed internet access via standard telephone lines, have again led to the development of new A/D converter families. This time, the requirements are resolutions of twelve to fourteen bit and conversion

rates of 1 MHz up to over 10 MHz, depending on the particular service. Moreover, the converters should be favorably priced and integration on the same chip as digital circuitry should be possible to enable a high level of integration.

With the invention of new technologies new challenges will appear that have to be solved by appropriate A/D converter designs. New integration technologies, on the other hand, will enable new converter architectures as well as the refinement of the existing ones.

1.3 Design Challenges

The most important goals in A/D converter design are conversion speed, resolution and linearity. While the resolution of an A/D converter basically describes the number of different output values the converter is able to generate, linearity is a measure of the quality of that A/D converter.

Especially for A/D converters with a resolution of twelve bit or more, achieving the necessary linearity is difficult. To improve the linearity, special post-processing techniques such as laser trimming can be used. But this involves additional manufacturing steps and is thus time-consuming and expensive. In addition, the thin film resistors that can be trimmed by laser are only available in special analog technologies. Therefore, a co-integration of the analog front-end and the digital signal processing circuitry is likely to be impossible.

Another method is to include self-calibration circuits on the chip [7–9]. This solution has the advantage that no special technology has to be used and no expensive post-processing steps are necessary. On the other hand, the additional circuitry adds to the complexity of the entire system, therefore the silicon area and the power consumption is increased. Moreover, the calibration action that has to be performed in fixed time intervals to cancel drift, e. g. due to changes in temperature, can interfere with the normal operation. To circumvent this problem,

special circuits have been developed that perform the calibration in the background [10–12].

Nevertheless, a solution without trimming or calibration is preferable. This means that the linearity must be achieved entirely by the matching capabilities of active or passive devices. The traditional solution is to use passive elements for matching, i. e. capacitor or resistor arrays. The matching accuracy that can be achieved with passive devices is usually limited to 0.1% or ten bit [13, 14], although better matching has been reported [15]. The use of resistors or capacitors, however, means to choose an analog-style process that provides a double-poly or equivalent option for high-precision capacitors or high-resistive polysilicon to be able to implement accurate resistors with typical values of several kilohms to several tens of kilohms. Although these options used to be available in processes with feature sizes around 1 µm, most deep submicron technologies are optimized for pure digital design and thus do not offer such luxury. Moreover, the silicidation step that is performed in deep submicron processes to decrease the sheet resistance of polysilicon makes the integration of resistors even more difficult.

If no passive elements are available, MOS transistors can be used as matching elements. The matching capabilities of MOS transistors have been studied extensively over the last 15 years [16–19]. Table 1.1 presents an overview of the results obtained in these publications. The numbers show that the mismatch in MOS transistors can be below 0.1% and is therefore in the same order of magnitude as the measured mismatch of passive devices. From this point of view, no advantage can be seen in using passive elements instead of MOS transistors. The processes used for these measurements, however, have big feature sizes compared to today's technologies.

The question therefore remains whether modern deep submicron processes are also capable of providing a matching accuracy comparable to the numbers reported above. It could be assumed that the matching improves with decreasing feature size because of improved lithography.

Table 1.1: Matching capabilities of MOS transistors [17–19]

Process	Measured	Result	Conditions
3.5 µm	current ratio of current mirrors	$\sigma < 0.2\%$	$W \times L >$ 25 µm × 50 µm
3 µm	current matching of MOSFETs	$\sigma_{V_T}, \sigma_\beta < 0.2\%$ $\sigma \approx 1\%$	W = 48 µm L = 12 µm
2.5 µm	single MOSFETs	$\sigma_{V_T} < 1$ mV $\sigma_\beta < 0.1\%$	$W, L >$ 20 µm

It has indeed been shown that the standard deviation of the threshold voltage decreases in modern submicron processes [20]. It will be one of the goals of this thesis to examine the matching capabilities of MOS transistors in a deep submicron technology.

As mentioned in the beginning of this section, achieving high speed is the second design goal of an A/D converter implementation. The speed can be increased by several measures. First, the improvements in process technologies and the trend towards smaller feature sizes enable the design of faster circuits. But even more important is the development of new architectures and gaining a deeper understanding of existing solutions to improve their performance.

A last challenge also has to be mentioned. Due to the trend towards system-on-a-chip where all components of a complete system are integrated on a single chip to decrease cost, size, weight and power consumption, A/D converters must be integrated using digital processes rather than technologies that are optimized for analog applications.

Especially deep submicron digital processes all have properties that are not appreciated in analog designs, except the speed: low power sup-

ply that decreases the dynamic range, no analog options as discussed above, and difficulties in achieving high gain. Nevertheless, when the appropriate architectures and design solutions are chosen, high-performance A/D converters can be created.

1.4 Organization of the Thesis

This thesis discusses the development and design of A/D converters using MOS transistors as matching elements in a special type of R-2R ladder structure, as mentioned at the beginning of this chapter. The MOSFET-only ladder yields a compact and versatile D/A converter which is used in two different types of A/D converter implementations. The structure of the thesis follows the development from the basic principle to the final converter.

Chapter 2 describes the MOSFET-only ladder. The underlying principle, the physical background as well as second-order effects are discussed. A thorough understanding of the second-order effects is necessary to be able to optimize designs for maximum accuracy. Measurements of bare MOSFET-only ladders are presented.

In chapter 3 this MOSFET-only ladder is used as a D/A converter in a successive-approximation A/D converter. One of the goals of this design is to show that the ladder structure is capable of achieving high accuracy not only statically, but also dynamically in an A/D converter. The resolution of this converter is 10 bit, the conversion speed is 200 ksample/s.

The good measurement results of the successive-approximation A/D converter allow to go one step further in the development, namely to increase the conversion speed. In chapter 4, several possible architectures are discussed, and finally a pipelined converter is chosen. The theory of pipelined operation is discussed, and an analysis of the errors that can occur in the individual building blocks of a pipeline stage and affect the linearity of the A/D converter is given.

The design of the pipelined 13 bit A/D converter is described in detail in chapter 5. The architecture as well as the influences the use of the MOSFET-only ladder has on the architecture of the pipeline are discussed. Design guidelines are given for the individual building blocks. A presentation of measurement results closes this chapter.

Chapter 6 begins with a summary of the thesis for the hurried reader. This summary also serves as the basis for concluding comments on the MOSFET-only ladder and the two A/D converter designs.

2

The MOSFET-Only Ladder

So stolz ich bin, muß ich mir selbst gestehn:
Dergleichen hab' ich nie gesehn.

MEPHISTOPHELES, Der Tragödie zweiter Teil

Linearity is one of the most important issues in A/D converter design if more than just a moderate resolution of about 8 bit is desired, as discussed in chapter 1. Therefore, special attention has to be paid to those building blocks that determine the linearity of an A/D converter.

In this chapter, the D/A converter which is used in both of the A/D converter designs described throughout this thesis, will be examined in detail. After presenting the functional principle of the main circuit of the D/A converter, the MOSFET-only R-2R-type ladder, the focus will be on second-order effects. These effects, although often

neglected, are important in achieving high accuracy. Therefore, it will be analyzed which second-order effects have an influence on the accuracy of the circuit, and which actions have to be taken to obtain the maximum linearity.

2.1 The Current Division Principle

The most important task that has to be performed inside a D/A converter is the accurate weighting of voltages, currents or charges [1,21,22]. In most of the cases binary weighting is necessary. Passive elements are normally used for this purpose, e. g. resistors for R-2R ladders and voltage-scaling D/A converters and capacitors for charge-scaling architectures.

It has been mentioned in chapter 1 that the exclusive use of MOS transistors is preferable over the employment of passive elements, especially when pure digital deep submicron processes are used for the implementation.

The traditional MOSFET-only architecture is a current-steering D/A converter using binary weighted current sources which are composed of identical unit MOS transistors. This type of D/A converter has been used in a wide range of applications [23–25]. The architecture allows the design of high-speed converters, but a severe drawback is the large silicon area it consumes due to the high number of unit current sources, which increases exponentially with the number of bits of the D/A converter. Moreover, the large area the current sources consume makes the matching of all MOS transistors in the array difficult. A sophisticated layout and a systematic distribution of the current sources forming the most significant bits is therefore necessary [24, 25]. Another solution would be some kind of self-calibration, such as dynamic element matching [26], but this adds further to area and power consumption.

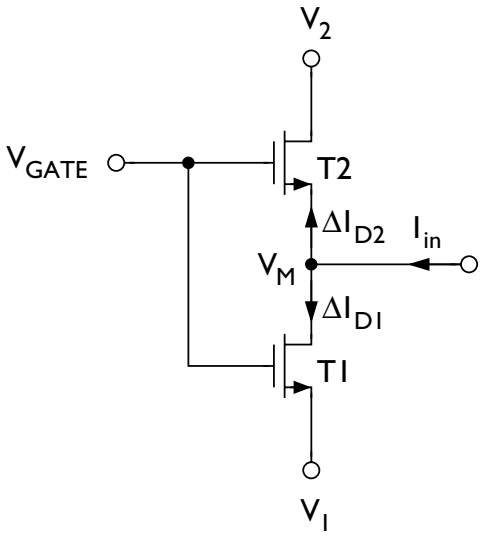

Figure 2.1: The basic current division principle

2.1.1 Current Division Using MOS Transistors

A different approach is using MOS transistors to perform current division resembling the principle of a resistive R-2R ladder. The fundamental difference between resistors and MOS transistors is the fact that in contrast to ideal resistors the relationship between current and voltage is non-linear in all operation regions of the MOSFET. Although a resistive ladder relies on the linearity of its passive elements, it can also be viewed as a device that subsequently divides the input current by a factor of two. This current division can be accomplished in an inherently linear way with MOS transistors despite their non-linear V–I characteristic [27]. The basic circuit is depicted in figure 2.1.

An input current I_{in} is divided into two currents ΔI_{D1} and ΔI_{D2}, respectively. The ratio of the two currents is given by

$$\frac{\Delta I_{\text{D1}}}{\Delta I_{\text{D2}}} = \frac{W_1/L_1}{W_2/L_2} \tag{2.1}$$

where W and L denote the dimensions of transistors *T1* and *T2*, respectively.

The ratio $\Delta I_{D1}/\Delta I_{D2}$ is constant and independent of I_{in}. It is also independent of the values of the terminal voltages V_1, V_2 and V_{GATE}. Moreover, the principle holds in all operation regions of the transistors *T1* and *T2*, regardless of being in weak, moderate or strong inversion. The current division is also independent of whether one or both transistors operate in saturation or are nonsaturated.

The current division principle is based on the symmetry of an MOS transistor with regard to its drain and source terminals. This symmetry can be seen by developing a general expression for the drain current which is valid in all operation regions. For this purpose, all terminal voltages of the transistor will be referenced to the bulk (substrate) voltage and not to the source voltage as it is preferred by circuit designers.

In general, the inversion layer current of an MOS transistor consists of a drift and a diffusion component. For an NMOS transistor, the inversion layer current at a certain position x in the channel (see figure 2.2) can be expressed as [27, 28]

$$I(x) = I_{drift}(x) + I_{diff}(x) \tag{2.2}$$

The current due to drift is caused by an electric field in the channel which can be expressed in terms of the surface potential $\psi_s(x)$. The drift component is thus

$$I_{drift}(x) = \mu W(-Q'_I)\frac{d\psi_s}{dx} \tag{2.3}$$

where μ denotes the surface mobility and Q'_I is the (negative) inversion layer charge per unit area.

The diffusion component of the inversion layer current can be evaluated as

$$I_{diff}(x) = \mu W \phi_t \frac{dQ'_I}{dx} \tag{2.4}$$

with ϕ_t being kT/q, the thermal voltage.

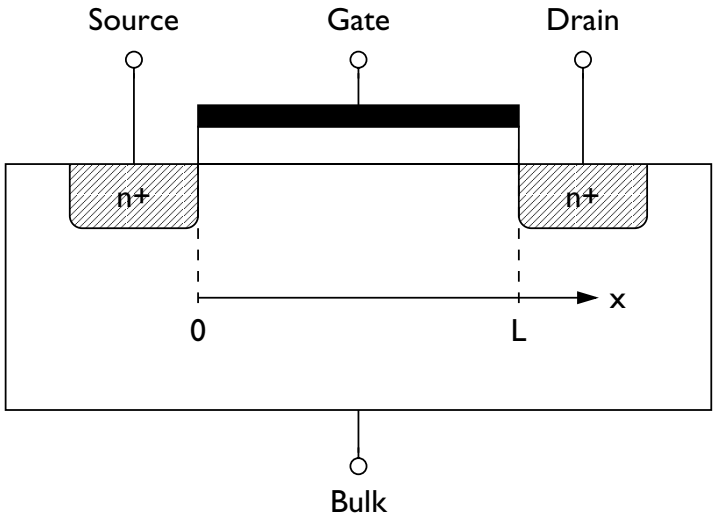

Figure 2.2: Cross-section of an NMOS transistor to calculate the drain current

In the steady state the current must be identical for all positions x in the channel and equal to the drain current I_{D}. Using equations (2.3) and (2.4) in (2.2) yields

$$I_{\mathrm{D}} = \mu W(-Q'_{\mathrm{I}})\frac{\mathrm{d}\psi_{\mathrm{s}}}{\mathrm{d}x} + \mu W \phi_{\mathrm{t}} \frac{\mathrm{d}Q'_{\mathrm{I}}}{\mathrm{d}x} \tag{2.5}$$

The integration of equation (2.5) leads to an expression for the drain current I_{D} which is valid in all operation regions:

$$I_{\mathrm{D}} = \frac{W}{L}\left[\int_{\psi_{\mathrm{s}0}}^{\psi_{\mathrm{s}L}} \mu(-Q'_{\mathrm{I}})\,\mathrm{d}\psi_{\mathrm{s}} + \phi_{\mathrm{t}} \int_{Q'_{\mathrm{I,source}}}^{Q'_{\mathrm{I,drain}}} \mu\,\mathrm{d}Q'_{\mathrm{I}}\right] \tag{2.6}$$

where $\psi_{\mathrm{s}0}$ and $\psi_{\mathrm{s}L}$ denote the surface potential at the source and drain end of the channel, respectively. Similarly, $Q'_{\mathrm{I,source}}$ and $Q'_{\mathrm{I,drain}}$ are the quantities of the inversion layer charge on the source and drain sides.

The inversion layer charge Q'_{I} can be expressed as a function of the terminal voltages as well as the surface potential [28]:

$$Q'_{\mathrm{I}} = -C'_{\mathrm{ox}}\left(V_{\mathrm{GB}} - V_{\mathrm{FB}} - \psi_{\mathrm{s}} - \gamma\sqrt{\psi_{\mathrm{s}}} \right) \tag{2.7}$$

with C'_{ox} denoting the gate oxide capacitance per unit area, V_{GB} is the gate-bulk voltage, V_{FB} denotes the flat-band voltage, and γ is the body-effect coefficient.

Equation (2.7) shows that the second integral in equation (2.6) is also a function of ψ_{s}. Therefore, the drain current can be written in the form

$$I_{\mathrm{D}} = \frac{W}{L}\left[f(\psi_{\mathrm{s}L}) - f(\psi_{\mathrm{s}0}) \right] \tag{2.8}$$

Equation (2.8) emphasizes the symmetry of the MOS transistor. If the drain and source terminals are interchanged, the only change will be the sign of the drain current I_{D}. The function $f(\psi_{\mathrm{s}})$ in equation (2.8) can be as complex as desired, including e. g. the body effect. Since the surface mobility μ is inside the integral in equation (2.6), all effects due to a non-constant mobility in the channel such as mobility degradation due to the normal field are covered as well.

The current division principle stated in equation (2.1) is a direct consequence of the symmetry shown in equation (2.8). Both transistors *T1* and *T2* in figure 2.1 share the identical gate-bulk voltage. Therefore, the function $f(\psi_{\mathrm{s}x})$ in equation (2.8) is identical for both transistors. Applying a current I_{in} in figure 2.1 changes the mid voltage V_{M}. This causes a change in $\psi_{\mathrm{s}L}$ for *T1* and $\psi_{\mathrm{s}0}$ for *T2* by the same amount. Therefore, the current changes $\Delta I_{\mathrm{D}1}$ and $\Delta I_{\mathrm{D}2}$ are identical for transistors with the same W/L ratio, otherwise $\Delta I_{\mathrm{D}1}$ and $\Delta I_{\mathrm{D}2}$ are scaled by the proportionality factors W_1/L_1 and W_2/L_2. This is exactly the expression found in equation (2.1). The mechanism derived above can also be visualized by means of a graphical MOS transistor model if numerical data to solve the integral in equation (2.6) is available [27, 29].

2.1.2 Limitations of the Current Division Principle

The current division is unaffected by all phenomena that can be modeled by equation (2.6). In particular, as discussed above, the body effect and mobility degradation due to the normal field in the MOSFET channel do not degrade the current division accuracy.

The so-called gradual channel approximation has been used to develop equations (2.2) to (2.8). In this approximation it is assumed that the lateral field in the channel is much smaller than the normal field, thus the direction of the field can be assumed to be perpendicular to the semiconductor surface. Therefore the calculations can be confined to the one-dimensional case.

In some circumstances, however, this one-dimensional approximation is not valid any more, and a two-dimensional field distribution has to be taken into account. This is especially the case for short-channel devices where practically the entire channel exhibits a two-dimensional electric field distribution.

For the current division principle, the discussion can be confined to relatively long channels because of matching considerations (see below). But even in this case, the one-dimensional channel approximation fails to give accurate results if the transistor starts to saturate and thus the field distribution on the drain side takes on a complicated two-dimensional shape [28, 30].

Therefore, drain-related effects such as channel-length modulation and drain-induced barrier lowering cannot be expected to be modeled adequately by equations (2.2) to (2.8).

A third effect that influences the accuracy of the current division is velocity saturation which will be examined in detail in section 2.2.

Mismatch will also cause deviations from the ideal current division in equation (2.1) (see chapter 1 for an overview of the matching accuracy of MOS transistors and passive elements). While mismatch in the transistor width W, length L and oxide thickness will only change the ratio of the current division (which otherwise stays ideal), mismatch in the threshold

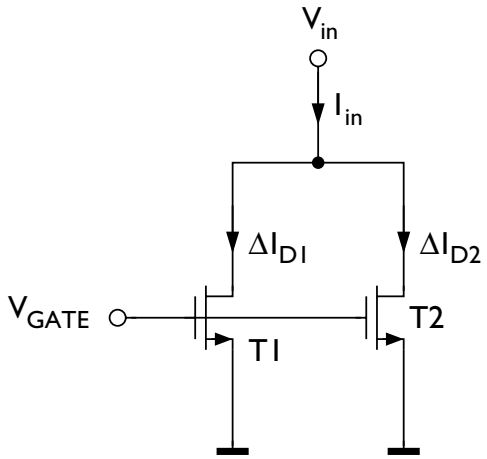

Figure 2.3: Preferred current division circuit

voltages of the transistors will not only alter the division ratio, but also renders the current division dependent of the input current, and thus introduces distortion [27].

2.1.3 Preferred Current Division Circuit

The discussion above leads to a circuit similar to the one depicted in figure 2.3 to be used for current division purposes. By proper choice of the terminal voltages and currents the transistors *T1* and *T2* operate in the linear region which reduces the drain-induced errors mentioned above.

Choosing a large effective gate voltage V_{GATE} lowers the current mismatch due to threshold voltage mismatch. Although a large gate overdrive is normally avoided to reduce non-linearity due to mobility degradation, this is not necessary in this case since the current division is independent of mobility variations as stated above. This is an interesting feature especially for deep submicron processes where the gate oxide becomes very thin and the normal field increases despite the efforts to realize constant-field scaling. Therefore, surface scattering and consequently mobility degradation becomes worse [31].

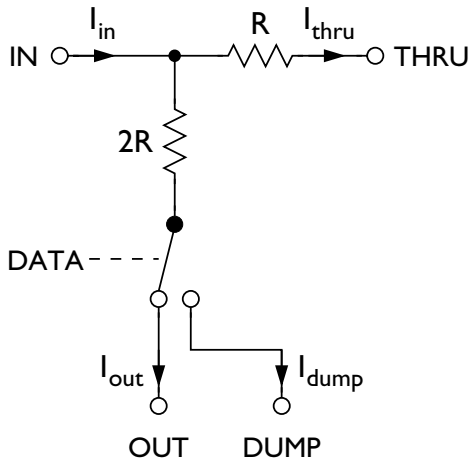

Figure 2.4: One slice of an R-2R ladder

Comparing figure 2.3 to figure 2.1, it can be seen that for both voltages V_1 and V_2 the same value has been selected (ground in the case of figure 2.3). This prevents any current other than I_{in} from flowing, which is advantageous from a power consumption point of view.

The circuit in figure 2.3 leads to an alternative way of looking at the current division principle. The two transistors *T1* and *T2* can be viewed as a resistive current divider with non-linear resistors. But this non-linearity does not compromise the accuracy of the current division as long as the non-linear V–I characteristics of the two devices are identical. This means that the accuracy only depends on the matching of the two devices but not on the linearity of their respective V–I curves.

2.2 The MOSFET-Only Ladder

Based on the circuit depicted in figure 2.3, an R-2R-like ladder can be built using MOS transistors only. Figure 2.4 shows one slice of a classical resistor-based ladder. The same functionality can be achieved

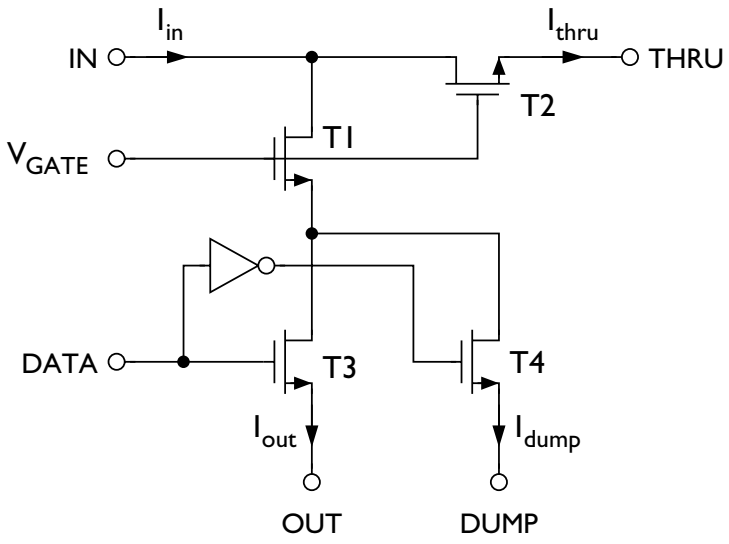

Figure 2.5: One slice of the actual MOSFET-only R-2R ladder

in a MOSFET-only implementation with the circuit shown in figure 2.5 [27, 32].

The R part of the original ladder slice is performed by transistor *T2*. The devices *T1* and *T3* (or *T4*) act as the $2R$ portion of the resistive counterpart. The incoming current I_{in} is divided into two equal currents (provided that the W/L ratio of all devices is identical), one of which leaves the slice as I_{thru}, the other one is switched either to I_{out} or I_{dump}. Transistors *T3* and *T4* therefore do not only act as a part of the resistive path, but also as switches. Thus, the entire slice consists of only four identical MOS transistors which leads to a very regular and compact layout which in turn is advantageous for good matching.

The current division principle stated in section 2.1 has been derived under the assumption that the gate-bulk voltage of the two devices is identical. Therefore it is essential for the circuit in figure 2.5 to function properly that the high-level of the signal DATA equals exactly the voltage V_{GATE}. To completely turn off the switch transistors, the low-level of DATA should be well below the threshold voltage of *T3* and *T4*.

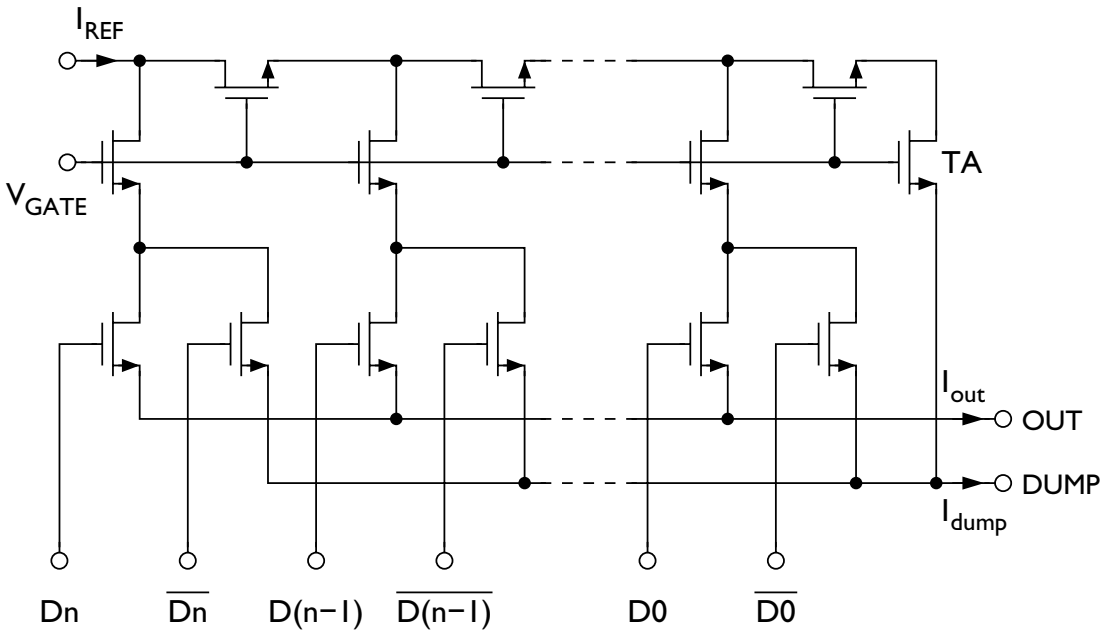

Figure 2.6: The complete MOSFET-only ladder

It should be pointed out that, although the structure of the MOSFET-only ladder is identical to the classical R-2R counterpart, the function is somewhat different. The small-signal resistances seen between the drain and source terminals of the individual transistors in the MOSFET ladder are not identical, nor do they have to be. The proper division of I_{in} only relies on the identical V–I characteristics of the transistors within the ladder.

Starting from one slice, it is straightforward to build a complete ladder circuit, which is depicted in figure 2.6. It consists of several slices in series and is terminated by transistor *TA*. The structure is biased by a reference current source I_{REF}, the value of which determines the full scale output current of the ladder.

As it is the case for a resistor-based R-2R structure, the entire ladder can be viewed as a sequence of MOS transistors in series and parallel, where each four transistors form a single MOSFET with unit W/L ra-

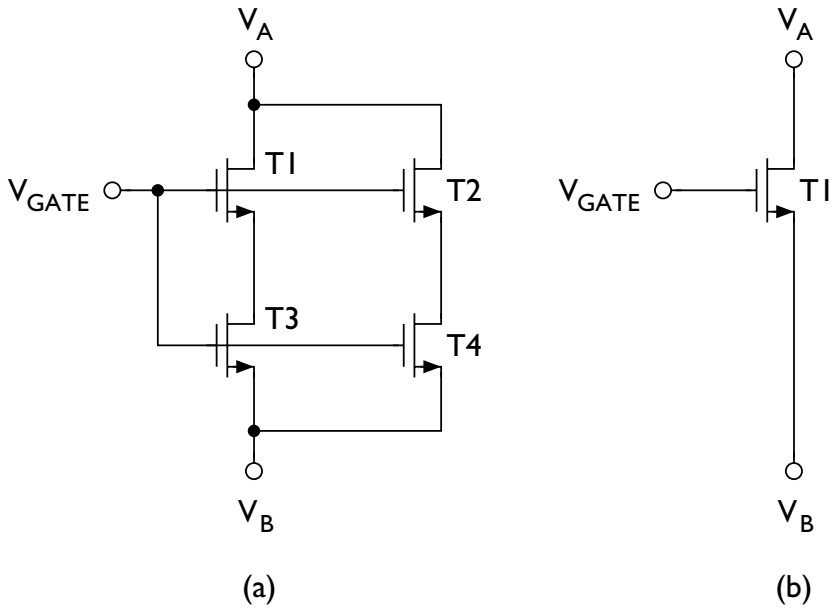

Figure 2.7: Equivalent circuit of four MOS transistors in the ladder

tio. In contrast to the resistive implementation, this series-parallel connection has additional consequences for the MOSFET-only ladder that have to be taken into consideration. The relevant parts of the circuit are shown in figure 2.7.

When comparing the schematic consisting of the four transistors *T1...T4* in figure 2.7(a) with the single equivalent MOS device in figure 2.7(b), it is evident that the drain-source voltages of transistors *T1...T4* on the left hand side are smaller than the voltage across the single device on the right. If their respective drain-source voltages differ, the lateral field $\mathcal{E}_x$ in the channel also differs. Although all transistors in the ladder operate in the linear region, the relationship between the lateral field $\mathcal{E}_x$ and the carrier velocity v_d is not exactly linear. The deviation from the linear relationship is small, however, for most of the devices, but has to be taken into account at least for the transistors on the MSB side of the ladder which normally operate in the vicinity of the saturation region. The non-linear relationship causes a change in the

V–I characteristic of an MOS transistor and can be taken into account as follows.

A simple model for velocity saturation can be given by [28]

$$|v_\mathrm{d}| = |v_\mathrm{d}|_\mathrm{max} \frac{|\mathcal{E}_x| / \mathcal{E}_\mathrm{c}}{1 + |\mathcal{E}_x| / \mathcal{E}_\mathrm{c}} \tag{2.9}$$

where $|v_\mathrm{d}|_\mathrm{max}$ is the maximum carrier velocity in the bulk material. $\mathcal{E}_\mathrm{c}$ denotes a critical value of the lateral field above which velocity saturation occurs.

Fortunately, it turns out to be easy to incorporate this effect into the description of the drain current of an MOS transistor. The drain current in the linear region including velocity saturation effects[1] can thus be written as [28]:

$$I_\mathrm{D} = \frac{1}{1 + (V_\mathrm{DS} / L\mathcal{E}_\mathrm{c})} \frac{W}{L} \mu_\mathrm{eff} C'_\mathrm{ox} \left[(V_\mathrm{GS} - V_\mathrm{T}) V_\mathrm{DS} - \frac{1}{2} V_\mathrm{DS}^2 \right] \tag{2.10}$$

Equation (2.10) makes evident that velocity saturation effects cause a difference in the V–I characteristics of the circuit in figure 2.7(a) and 2.7(b). Because the current division relies on identical V–I characteristics, velocity saturation will deteriorate its accuracy. Since the drain-source voltages change throughout the ladder, the change in the V–I curve differs for every MOS device in the ladder. This makes it difficult to completely cancel this non-ideality. Moreover, predicting the value of the critical electric field $\mathcal{E}_\mathrm{c}$ is difficult. In the literature, values between 7 kV/cm and 60 kV/cm have been reported [28, 33–36].

1 It might be noted that the term *velocity saturation effect* is used in a very general manner here to refer to any non-linear relationship between v_d and $\mathcal{E}_x$. It does not necessarily mean that the lateral field is large enough to actually cause saturation of v_d.

The negative influence of the velocity saturation effects can be decreased sufficiently by choosing a large channel length. This can be seen by rewriting equation (2.10):

$$I_D = \frac{W}{L_{eff}} \mu_{eff} C'_{ox} \left[(V_{GS} - V_T) V_{DS} - \frac{1}{2} V_{DS}^2 \right] \tag{2.11}$$

with

$$L_{eff} = L + \frac{V_{DS}}{\mathcal{E}_c} \tag{2.12}$$

which shows that the slight changes in the drain current which are due to velocity saturation effects can be viewed as if they were caused by a small deviation of the transistor length from its ideal value. Therefore, to keep the error due to velocity saturation effects low, it has to be guaranteed that the rightmost factor of L_{eff} in equation (2.12) is much smaller than the ideal length L.

2.3 Offset Sensitivity of the Ladder

Another problem that can cause severe errors in the output current I_{out} is an offset voltage between the two output terminals OUT and DUMP in figure 2.6. A ladder circuit only functions accurately if both output nodes are at exactly the same potential. Only if this condition is satisfied, the ladder truly consists of devices in series and in parallel, which is necessary to obtain binary weighted outputs. Any offset between the terminals violates this condition, thus leading to an error in the output current. The offset-induced error depends on the output resistance of the ladder which changes with respect to the digital word applied to the ladder. Therefore, this error is signal-dependent and causes distortion of the output signal. To calculate the error, a general description of the output resistance is necessary. For simplicity, this description will be developed for a resistive ladder.

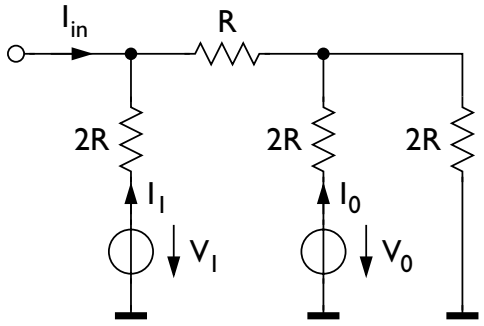

Figure 2.8: Calculating the output resistance of an R-2R ladder

2.3.1 Output Resistance of an R-2R Ladder

2.3.1.1 A Two Bit Ladder

The output resistance of a ladder, i. e. the resistance seen when looking inside one of the terminals OUT or DUMP in figure 2.6, depends on the digital word applied to the ladder, as discussed before. Therefore, a closed expression for all digital codes will be developed. A two bit ladder will be considered as an example before expanding the result to an arbitrary number of bits. This two bit ladder is depicted in figure 2.8. The ladder is assumed to be biased by a reference current source I_{in} with infinite output resistance.

Instead of the switches and output terminals OUT and DUMP, as in figure 2.6, all branches are terminated by voltage sources of identical value (V_0 and V_1 in figure 2.8). The output resistance is calculated by determining the currents I_0 and I_1 caused by the voltage sources. In this way the output resistance for e. g. terminal OUT for a particular code can be calculated by summing up the appropriate currents. Only the output resistance of one terminal will be calculated. The value for the other terminal is identical due to the symmetry of the ladder.

The following notation will be used. I_{jk} denotes the part of current I_j which is caused by voltage source V_k. Superposition then yields the complete current I_j.

To start with, only voltage source V_1 is considered. The two currents can be evaluated as

$$I_{11} = \frac{V_1}{4R} \qquad I_{01} = \frac{-V_1}{8R} \tag{2.13}$$

R denotes the unit resistance of the ladder. Repeating the calculation, this time considering only voltage source V_0, yields the remaining two currents.

$$I_{00} = \frac{V_0}{\frac{16}{5}R} \qquad I_{10} = \frac{-V_0}{8R} \tag{2.14}$$

With these four currents the output resistance for all codes can be computed. The complete branch currents I_1 and I_0 can be determined as follows:

$$I_1 = b_1(b_1 I_{11} + b_0 I_{10}) \tag{2.15}$$

$$I_0 = b_0(b_1 I_{01} + b_0 I_{00}) \tag{2.16}$$

where b_i denotes bit i of the digital code applied to the ladder with b_0 being the LSB. If the bit is set, b_i equals one, otherwise it is zero.

Starting from the branch currents, the calculation of the output resistance for individual codes is straightforward. Code 01, e. g., yields an output resistance of $3.2\,R$. The result is the same for code 11 due to the symmetry of the ladder. The maximum output resistance equals $4\,R$ and is obtained at code 10.

2.3.1.2 General Calculation

To extend the calculation to an arbitrary number of bits, it is convenient to examine the resistance of an unterminated ladder first. This special circuit is shown in figure 2.9. The termination resistor on the right hand side is missing, therefore the ladder cannot be reduced to a single R as would be the case for a complete R-2R ladder. This unterminated circuit occurs at the MSB side of a conventional ladder when

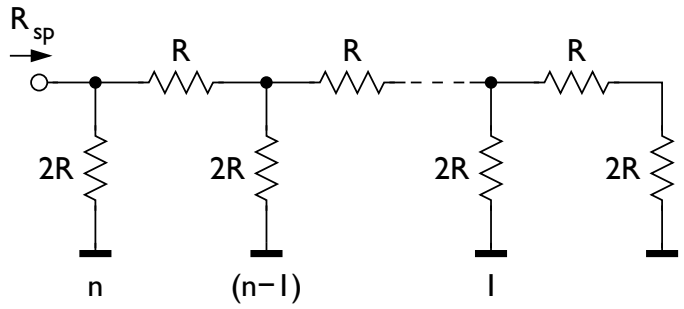

Figure 2.9: Unterminated R-2R ladder

calculating the output resistance. The calculation of the resistance R_{sp} is straightforward and yields

$$R_{\mathrm{sp}}(n) = \frac{\left[\sum_{i=0}^{n} 2^{2i}\right] + 1}{\sum_{i=0}^{n} 2^{2i}} R \tag{2.17}$$

where n denotes the number of bits of the unterminated ladder part. The equation also holds for the case with zero bits where only the rightmost resistor in figure 2.9 is present.

Now the ground has been prepared to calculate the general case. As for the two bit ladder examined above, expressions for the currents I_{jk} will be needed. In general, these currents are of the following form:

$$I_{jk} = \alpha_{jk} V_k \frac{1}{R_{jk}} \tag{2.18}$$

where α_{jk} is defined as follows:

$$\alpha_{jk} = \begin{cases} 1 & \text{if } j = k \\ -1 & \text{if } j \neq k \end{cases} \tag{2.19}$$

The resistances R_{jk} can be calculated using the values of $R_{\mathrm{sp}}(n)$ defined

in equation (2.17). It is worth noting that, due to the symmetry of the ladder, the following relationship holds:

$$R_{jk} = R_{kj} \qquad \forall\, j, k \tag{2.20}$$

Therefore, only about half of the resistance values have to be determined. The calculation is lengthy but straightforward and yields

$$R_{jk} = \begin{cases} \dfrac{2^{(k-j+1)} R \left[R_{\mathrm{sp}}(n-1-k) + 2R \right]}{R_{\mathrm{sp}}(n-1-k)} & \text{for } j < k \\ R_{\mathrm{sp}}(n-1-k) + 2R & \text{for } j = k \\ R_{kj} & \text{for } j > k \end{cases} \tag{2.21}$$

where n denotes the number of bits of the entire R-2R ladder.

The next step towards the result is the summation of the individual currents of one branch. The general branch currents I_j can be expressed as

$$I_j = b_j V_{\mathrm{in}} \sum_{i=0}^{n-1} b_i \alpha_{ji} \frac{1}{R_{ji}} \tag{2.22}$$

In this equation the various voltage sources V_k of equation (2.18) have been replaced by a single source V_{in} which is possible because their voltages are identical.

To finally obtain an expression for the output resistance of the R-2R ladder, all branch currents have to be summed up, the resulting current must be divided by the input voltage V_{in}, and the reciprocal value has to be taken. These operations lead to the following general expression for an n bit R-2R ladder:

$$R_{\mathrm{out}}(b_{(n-1)\ldots 0}) = \frac{1}{\sum_{j=0}^{n-1} b_j \sum_{i=0}^{n-1} b_i \dfrac{\alpha_{ji}}{R_{ji}}} \tag{2.23}$$

2.3.1.3 Minimum Output Resistance

To calculate the output current error due to offset, the minimum and maximum output resistance of the ladder must be known. Finding the upper limit presents no problem. Ignoring the case of code 00...0 where no branch of the ladder is connected to the output and therefore the output resistance seen at terminal OUT is infinite, the maximum occurs at code 10...0 when only the MSB is set. In that case the output resistance equals $4R$.

The case of minimum resistance is more complicated. Instead of presenting a formula for the minimum output resistance in closed form, the code where the minimum output resistance occurs is calculated. This code can then be used in equation (2.23) to determine the output resistance. With this procedure, more insight in the behavior of the ladder is gained. The second advantage is that the resulting code is also valid for the MOSFET ladder, while the output resistance of the resistive and the MOSFET-only implementation naturally will differ.

Due to the symmetrical nature of the ladder, the values of the output resistance are also symmetrical with respect to code 10...0. This means that the output resistance for code 01...11 is the same as for code 10...01 etc. Therefore, the minimum output resistance occurs at exactly two codes.

For an n bit ladder, the minima occur at codes s_{n_1} and s_{n_2}, where it is assumed that s_{n_1} is the lower code. For a ladder with $(n+1)$ bits, the minima occur at codes $s_{(n+1)_1}$ and $s_{(n+1)_2}$. The lower code has its MSB set to zero, which effectively makes the $(n+1)$ bit ladder behave as an n bit ladder. Thus the following holds true for arbitrary values of n:

$$s_{(n+1)_1} = s_{n_2} \tag{2.24}$$

or in other words, the lower code of the $(n+1)$ bit ladder is the same as the higher code of the n bit ladder.

For an n bit ladder, the minimum output resistance therefore occurs at code

$$s_n = \sum_{i=0}^{n-1} (-1)^{(i+n-1)}\, 2^i \tag{2.25}$$

where s_n is the lower one of the two codes (s_{n_1} in the discussion above). Equation (2.25) together with (2.23) thus allows to calculate the minimum output resistance of an arbitrary R-2R ladder.

2.3.1.4 MOSFET-Only Ladders

The calculations outlined above in principle also hold for the MOSFET-only ladder. The big difference, however, is that no equivalent to the unit resistance R of a classical resistive R-2R ladder exists (see the discussion in section 2.2).

The codes at which the maximum occurs are identical to the R-2R ladder and are thus predicted by equation (2.25). The voltage drop over the entire ladder divided by the reference current can be used instead of the value of R in the equations above.

This approximation of the unit resistance, however, leads to an estimation of the minimum output resistance which is too high. The reason is the decreasing slope of the V–I characteristic of an MOS transistor in the transition region between linear and saturated operation. For MOSFET ladders which are biased such that all transistors operate deeply in the linear region, equation (2.23) predicts the minimum resistance accurately. For optimized designs, however, where the transistors at the MSB side operate in the vicinity of saturation, the minimum output resistance is about 75% of the value predicted by equation (2.23).

2.3.2 Offset Sensitivity

After having developed a general expression for the output resistance of an R-2R ladder, the offset sensitivity can now be determined. Fig-

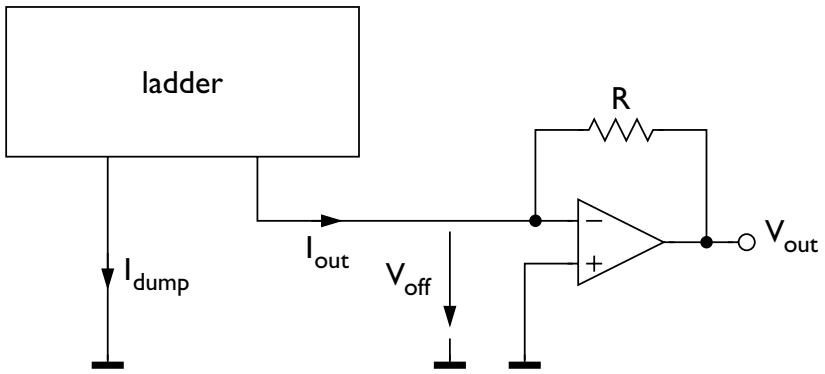

Figure 2.10: Offset sensitivity of the ladder

ure 2.10 shows the traditional situation when a transresistance based on an operational amplifier is used to convert the ladder output current I_{out} to an output voltage V_{out}.

Any offset voltage V_{off} of the operational amplifier will cause an error in the output current of the ladder, as was discussed at the beginning of this section. In general, this current error is [37]

$$\Delta I = \frac{V_{\text{off}}}{R_{\text{out}}} \tag{2.26}$$

where ΔI is defined as the difference between the real and the ideal output current of the ladder and R_{out} is the output resistance as given by equation (2.23).

The output resistance depends on the digital code applied to the ladder and is thus signal-dependent. This fact in turn renders the current error ΔI signal-dependent which is similar to distorting the output signal of the ladder.

The output error is bounded by

$$\frac{|V_{\text{off}}|}{4R} \leq \Delta I \leq \frac{|V_{\text{off}}|}{R_{\text{min}}} \tag{2.27}$$

with R_{min} being the minimum output resistance of the ladder as defined by equations (2.23) and (2.25). Equation (2.27) can also be viewed as

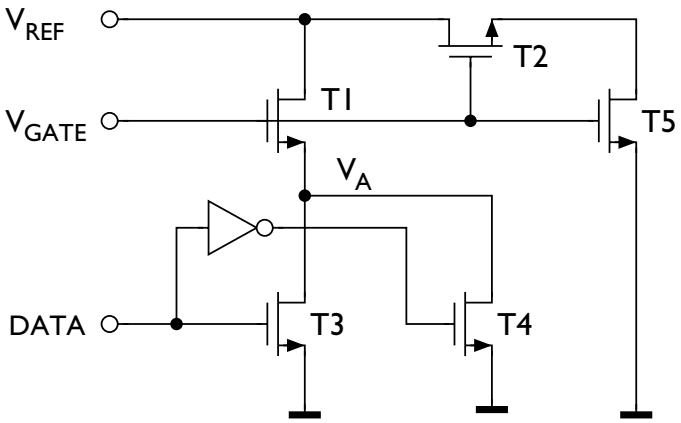

Figure 2.11: One bit ladder to calculate the output current settling

a bound for the maximum offset voltage of the operational amplifier when a maximum tolerable current error is given.

In chapter 3 a successive-approximation A/D converter will be presented. This circuit includes an amplifier with externally adjustable offset voltage. Therefore, offset can be applied to the ladder and the consequences on linearity can be studied. Measurements of the respective output of the spectrum will be presented in section 3.6.

2.4 Speed of the Ladder

The second important issue in an A/D converter design besides accuracy is speed. After having investigated several issues affecting the accuracy of the MOSFET-only ladder in the last sections, the dynamic behavior will be examined below.

The speed of the MOSFET-only ladder is determined by the equivalent resistance of the MOS transistors and their parasitic capacitances. A first-order calculation of the output current settling of the ladder can be carried out using a one bit ladder as shown in figure 2.11 [37].

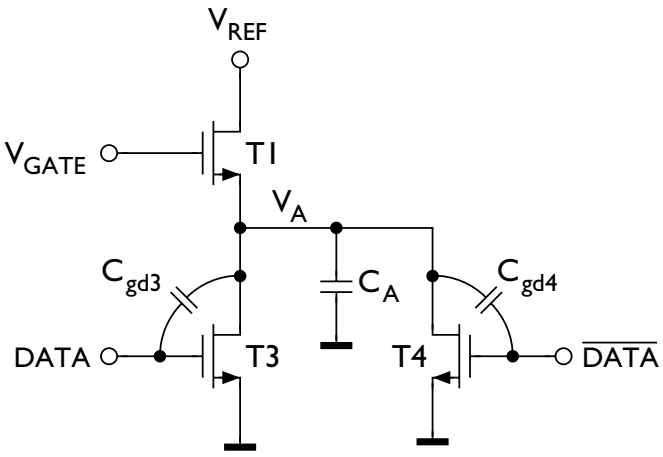

Figure 2.12: Simplified schematic to calculate the settling behavior of the ladder

Although the ladder inside a D/A converter is normally biased using a reference current source, it is more convenient for the following calculation to use a reference voltage at the ladder input. The dynamic behavior of the circuit is unaltered by this measure. All transistors are assumed to operate in the linear region, which is the case in almost any application. It is also assumed that the current is switched from transistor *T4* to *T3*. The settling behavior can then be determined by calculating the output response when switching the signal DATA from its low-level to the high-level.

In a first-order calculation, only the branch consisting of *T1*, *T3* and *T4* has to be considered, as depicted in figure 2.12. Three capacitances are of importance. C_A denotes the sum of all parasitic capacitances at node V_A and consists of the gate-source and source-bulk capacitances of *T1* as well as the drain-bulk capacitances of *T3* and *T4*, respectively:

$$C_A = C_{gs1} + C_{sb1} + C_{db3} + C_{db4} \tag{2.28}$$

The gate-drain capacitances of *T3* and *T4* cause capacitive coupling of the digital input signal to node V_A, which in turn disturbs the output

current of the ladder slice. Only by the time the voltage V_A has settled, the output current is stable.

Although the two capacitances C_{gd3} and C_{gd4} are driven by differential signals, their effects do not cancel completely. The rising and the falling edges of the signal DATA are never completely symmetrical, and a small delay always exists between the inverted and non-inverted signal. But more important, the two capacitances are also not identical. Transistor *T3* is in the off-state when the input signal is at its low-level. Therefore its gate-drain capacitance only consists of the overlap capacitance between gate and drain. When *T3* gets switched on by the rising edge of DATA, the channel has to be established before C_{gd3} reaches its final value. Although the time constant for the creation of the channel is small, this short-term difference between C_{gd3} and C_{gd4} is sufficient to disturb the voltage at node V_A.

The settling behavior of the ladder output current after switching the signal DATA from its low-level to the high-level is given by

$$\Delta I(t) = I_{sw} \frac{C_{gd3}}{C_A + C_{gd3} + C_{gd4}} e^{-t/\tau} \tag{2.29}$$

where $\Delta I(t)$ describes the deviation of the ladder output current from its final value after having switched the output current to *T3*. I_{sw} denotes the value of the current peak which depends on the asymmetry of the inverted and non-inverted digital signal on the one hand and on the behavior of the gate-drain capacitances of transistors *T3* and *T4* on the other hand, as discussed above. The time constant τ equals

$$\tau = \frac{r_1 r_3}{r_1 + r_3} (C_A + C_{gd3} + C_{gd4}) \tag{2.30}$$

where r_1 and r_3 denote the small-signal equivalent resistances between drain and source of *T1* and *T3*, respectively.

Capacitance C_A has a twofold effect on the settling. If C_A is increased, e. g. by adding extra capacitance to node V_A, the time constant increases according to equation (2.30). The peak value of the

settling component, on the other hand, decreases, as can be seen by equation (2.29). If the settling component could be attenuated to an extent that it would not affect the accuracy of the output current any more, the speed of the ladder could be increased significantly. Unfortunately, for most implementations the additional capacitance that would be needed is too large to be integrated. The four bit ladder used in the pipelined A/D converter described in chapter 5, e. g., would require an additional capacitance of more than 100 pF for each ladder slice, which is clearly infeasible in most processes.

To identify the parameters that affect the ladder speed, equation (2.30) can be examined further. Using the expression of C_A given in equation (2.28) in (2.30), the time constant can be written as

$$\tau = \frac{r_1 r_3}{r_1 + r_3} (C_{gs1} + C_{sb1} + C_{db3} + C_{gd3} + C_{db4} + C_{gd4}) \qquad (2.31)$$

The small-signal drain-source resistance of an MOS transistor in the linear region for small values of V_{DS} is

$$r = \frac{1}{\frac{W}{L}\mu_{eff} C'_{ox}(V_{GS} - V_T)} \qquad (2.32)$$

The gate-source capacitance is given by

$$C_{gs} = \frac{1}{2} W L C'_{ox} \qquad (2.33)$$

as long as the transistor is in the linear region, and the source-bulk capacitance equals

$$C_{sb} = C_J A_S + C_{JSW} P_S \qquad (2.34)$$

where C_J denotes the bulk junction capacitance, C_{JSW} is the sidewall bulk junction capacitance, and A_S and P_S denote the area and perimeter of the source, respectively. With the area and perimeter of the drain inserted, this equation also holds for the drain-bulk capacitance.

The gate-drain capacitance needs more attention. During settling of the output current, either transistor *T3* or *T4* is in linear region, the other one is switched off. The gate-drain capacitance is different in both cases and equals

$$C_{\mathrm{gd}} = \begin{cases} \frac{1}{2}\, W\, L\, C'_{\mathrm{ox}} & \text{in linear region} \\ C_{\mathrm{GDO}}\, W & \text{in off-state} \end{cases} \tag{2.35}$$

where C_{GDO} denotes the gate-drain overlap capacitance.

Using equations (2.32), (2.33), (2.34) and (2.35) in (2.31) results in the following expression for τ where it is assumed that all transistors have identical dimensions.

$$\tau = \frac{1}{\mu_{\mathrm{eff}}\, C'_{\mathrm{ox}}} \frac{1}{2V_{\mathrm{GATE}} - V_{\mathrm{T1}} - V_{\mathrm{T3}} - V_{\mathrm{A}}} \cdot \left[C'_{\mathrm{ox}} L^2 + (3C_{\mathrm{J}} l_{\mathrm{d}} + 6C_{\mathrm{JSW}} + C_{\mathrm{GDO}})L + 6\, C_{\mathrm{JSW}} l_{\mathrm{d}} \frac{L}{W} \right] \tag{2.36}$$

V_{GATE} denotes the gate voltage applied to the ladder, while V_{A} is the voltage at the midpoint of the ladder slice (see figure 2.11). l_{d} denotes the length of the diffusion area for the source and drain regions.

The rightmost factor in equation (2.36) is small compared to the other factors for practical designs of the ladder and can thus be omitted. Therefore, it is possible to write equation (2.36) in the following form:

$$\tau = \zeta_1 L^2 + \zeta_2 L \tag{2.37}$$

where ζ_1 and ζ_2 are proportional factors which depend on process parameters and the terminal voltages.

The strong dependence of the ladder speed on the channel length L is evident. Equation (2.37) illustrates the common compromise between accuracy and speed in A/D converter design. To achieve high speed, the channel length of the MOS transistors has to be made as small as

possible. Matching and velocity saturation effects, on the other hand, demand a large channel length.

Equation (2.37) also states that the ladder speed is independent of the value of the reference current of the ladder. If, e. g., for a given transistor length L the reference current is doubled, the width of all ladder transistors also has to be doubled to keep the voltage across the ladder unchanged. But according to equation (2.37) the settling time of the output current only depends on the transistor length, but not on the width. This can also be seen by noting that the necessary increase of the transistor width due to the larger current not only increases the parasitic capacitances of all transistors, but also decreases the small-signal drain-source resistance of the transistors by the same amount.

2.5 Practical Ladder Implementation

2.5.1 Transistor Dimensions

The theory and calculations presented in the last sections form the basis to determine the proper dimensions of the MOS transistors in the ladder.

The first quantity that has to be selected is the ladder reference current I_{REF}. The minimum value of I_{REF} is determined by the LSB current of the ladder, which has to be above the leakage current of the junction diodes of the drain and source areas, preferably by at least one order of magnitude.[2] The maximum value of the reference current depends on the power consumption that is tolerable.

2 For D/A converters used in subranging or pipelined A/D converters (see chapter 4), it is not the LSB current, which is of importance, but the effective resolution the D/A converter has to deliver. Especially in a pipelined A/D converter, the number of bits of the D/A converter is small. The resolution, however, has to be the full resolution of the A/D converter.

As pointed out in section 2.4, the speed of the ladder is, in a first-order approximation, independent of the reference current. Therefore, the speed requirements of a particular implementation cannot be used to determine the value of the reference current. Its choice, within the boundaries given above, largely depends on the requirements of the subsequent circuits that process the ladder output current.

Once the reference current is fixed, the transistor dimensions can be determined. The power supply of the circuit as well as constraints on the voltage at the output of the ladder allow a certain voltage swing over the entire ladder. For a given current, this voltage swing determines the W/L ratio of the MOS transistors.

The selection of the proper transistor length L needs more attention. Matching requirements pose a lower limit on the channel length. More accurately, the gate area is the critical value, for both important matching parameters, the variance of the threshold voltage $\sigma^2(V_T)$ and the variance of the conductance parameter $\sigma^2(\mu_{eff}C'_{ox})$ decrease with increasing gate area WL [17–20].

The second constraint which has to be obeyed are errors due to velocity saturation effects. This constraint also presents a lower bound on the channel length, although the resulting errors are hard to predict exactly, as was discussed in detail in section 2.2. Finally, the upper limit of the channel length is determined by the required speed of the ladder. This is illustrated by equation (2.37).

The discussion above shows that the choice of the transistor length of the MOS devices determines both the accuracy and the speed of the ladder. This single value represents the entire trade-off between precision and speed that characterizes the entire A/D converter design process.

2.5.2 Layout

For precise analog circuits the layout is of the same importance for the final accuracy as is the design of the circuit itself. This is especially true for circuits which require the best possible matching as the MOSFET-

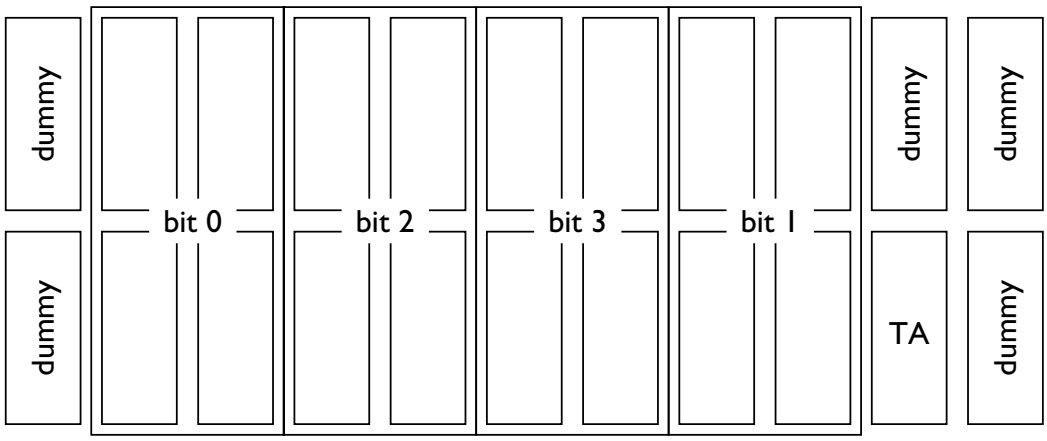

Figure 2.13: Layout of a four bit MOSFET-only ladder

only ladder. The structure of one slice of the ladder as depicted in figure 2.5 lends itself very well to a highly regular and compact layout.

To reduce errors due to major carry transitions, a one-dimensional common-centroid structure where the MSB slice is in the center of the ladder is used, which is illustrated in figure 2.13. Dummy structures at both ends of the ladder guarantee an identical environment for all active transistors.

Special attention has to be paid to the interconnect resistances in the ladder. Although any resistance which is common to all transistors, e. g. the contact resistance of the drain and source connections, does not compromise the accuracy of the ladder, any difference in these resistances does. It is obvious from figure 2.13 that it is impossible to match the length of the metal connections between the slices when using a common-centroid layout. So it is necessary to adjust the width of the metal wires as well as the number of metal layers in parallel (a luxury which only can be exploited with modern processes with a substantial number of metal layers) to match the interconnect resistance as closely as possible.

2.5.3 Output Impedance

One additional point which has to be taken into consideration when designing a MOSFET-only ladder is the output impedance which shows resistive as well as capacitive behavior. The resistive part has been covered in detail in section 2.3.1. Because it determines the offset sensitivity and thus the accuracy of the ladder, the output resistance has to be taken into account early in the design process. The output capacitance, on the other hand, does not affect the accuracy of the output current. It can, however, affect the circuit extracting the ladder current.

The output impedance of the ladder can be approximated by a first-order R–C network with one dominant pole. This pole depends on the digital code applied to the ladder. The lowest frequency occurs when all bits are switched to the output. The four bit ladder used in the pipelined A/D converter described in chapter 5 has a dominant pole located at a frequency of 3.5 MHz for the code 11...1, which, in conjunction with an output resistance of 2 kΩ, yields an equivalent output capacitance of more than 22 pF.

This large output capacitance results from the summation of individual parasitic capacitances which are present at each node of the ladder. The mechanism can be studied by examining a single branch of a ladder, which is shown in figure 2.14(a). The small-signal equivalent circuit is depicted in figure 2.14(b). All parasitic capacitances per node are lumped into one single equivalent capacitance. Although the transistors operate in the linear region, the full small signal circuit including the transconductance g_m and the output resistance r is used [21].

In a first step, only capacitance C_2 is considered, all other capacitances are ignored. The output impedance Z_{out} can be calculated as

$$Z_{out} = \frac{1 + sC_2r_3}{sC_2(1 + r_3g_{m3})} \tag{2.38}$$

It can be seen that the capacitance C_2 appears at the output, multiplied

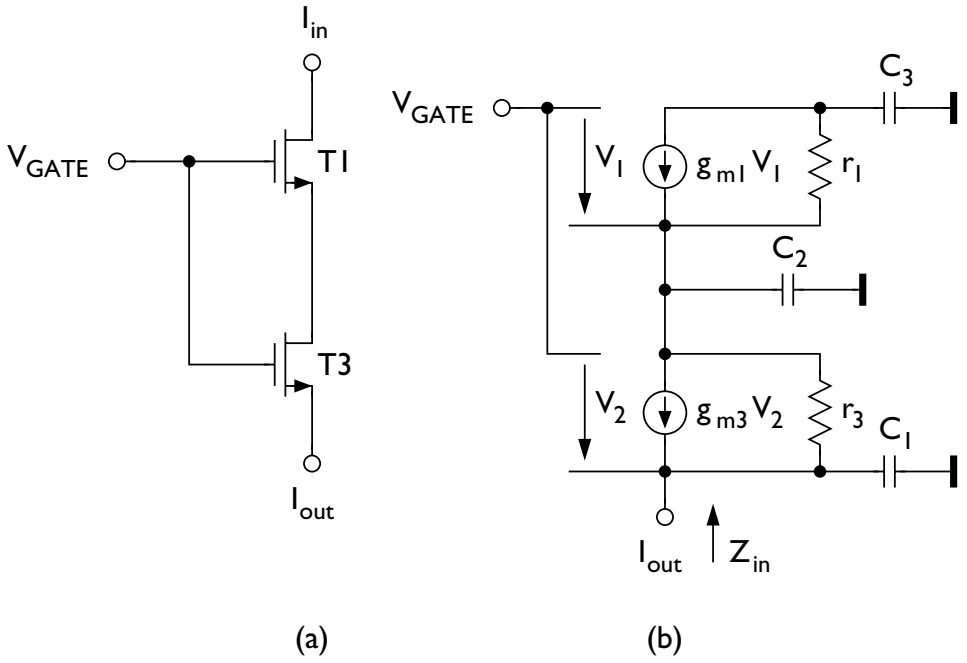

Figure 2.14: One branch of a MOSFET-only ladder (a) and its small-signal equivalent circuit (b)

by a factor $(1 + r_3 g_{m3})$. Although the value of $r_3 g_{m3}$, which equals the gain of the transistor, is normally smaller than unity in the linear region, it nevertheless enlarges the effect of the capacitance.

Carrying out the same calculation for capacitance C_3 and ignoring all other capacitances yields

$$Z_{out} = \frac{1 + sC_3(r_1 + r_3 + r_1 r_3 g_{m1})}{sC_3(1 + r_1 g_{m1})(1 + r_3 g_{m3})} \tag{2.39}$$

The result resembles the one before, only this time both transistors are involved in enlarging the capacitance.

Equations (2.38) and (2.39) show the mechanism how the output capacitance of the MOSFET-only ladder originates. The individual node

capacitances are summed at the output, multiplied by the gain of the transistors forming the path to the output.

If the output current of the ladder is summed at a low-impedance node, such as the input node of the regulated cascode described in section 3.3, the output capacitance has no noticeable effect on the settling of the circuit. But if the input impedance cannot be kept at a low level in all clock phases, as is the case during the auto-zero cycle of the summation circuit discussed in section 5.2.2, a low-frequency pole is created by the output capacitance of the ladder which significantly reduces the settling speed. In this case, special countermeasures such as switching off the ladder current at one side are necessary.

2.6 Ladder Measurements

Before starting the design of an A/D converter exploiting the MOSFET-only ladder as critical analog building block that determines the accuracy of the entire converter, extensive measurements on MOSFET-only ladders have been carried out [38].

Special care has been taken in the measurement setup to ensure that external effects which affect the measurement accuracy such as wire resistances are eliminated and the results reflect the true accuracy of the ladder.

A 1 μm digital CMOS process with two layers of metal was used for the measurements. The layout of the ladder follows the guidelines presented in section 2.5, that is compact layout, a one-dimensional common-centroid structure and dummy devices at both ends of the ladder.

A large channel length of 60 μm was chosen to minimize errors due to velocity saturation effects and to obtain a large gate area for good matching. The transistor width is 20 μm. The reference current of the

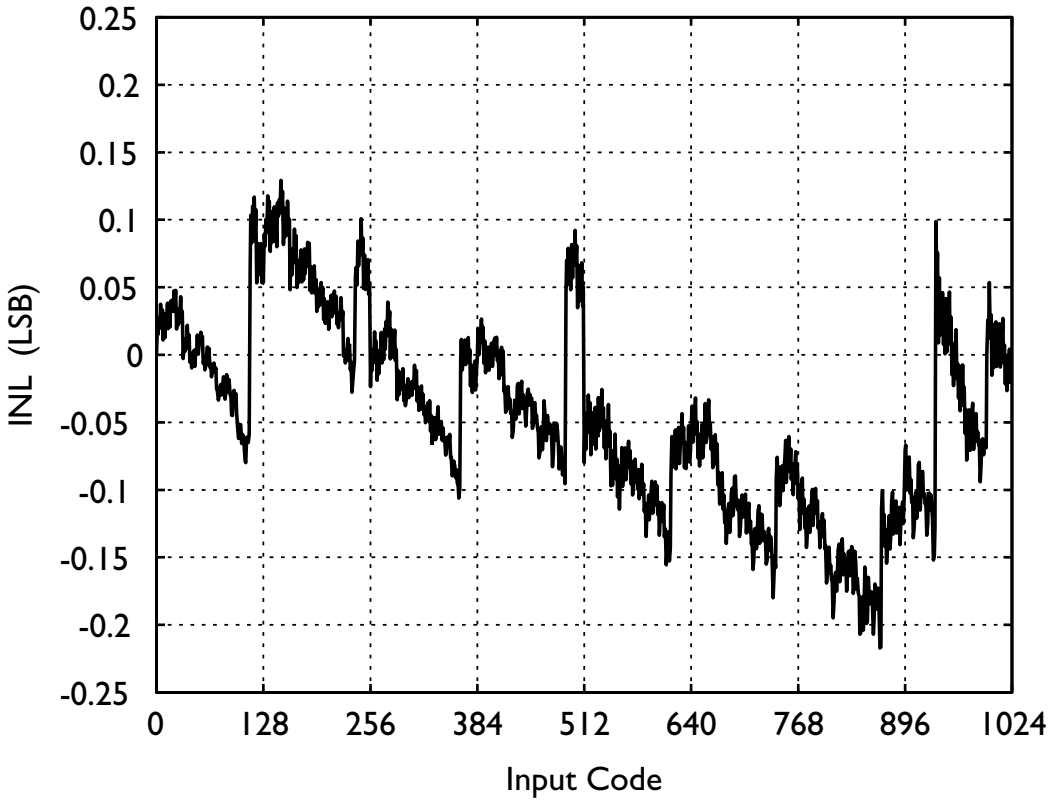

Figure 2.15: Measured static INL of a 10 bit ladder in a 1 µm CMOS process

ladder is 100 µA. This relatively low value was chosen in view of the power consumption of an A/D converter using this ladder.

The measured integral non-linearity (INL) of a typical 10 bit ladder is shown in figure 2.15. Different limits can be used to decide whether the linearity of an A/D or D/A converter matches its resolution. One practical limit is ±0.5 LSB for the INL which guarantees monotonicity of the converter [39]. In figure 2.15, the INL is below ±0.25 LSB for all codes, which means that a linearity better than 11 bit is achieved. The differential non-linearity (DNL) of the same ladder is depicted in figure 2.16.

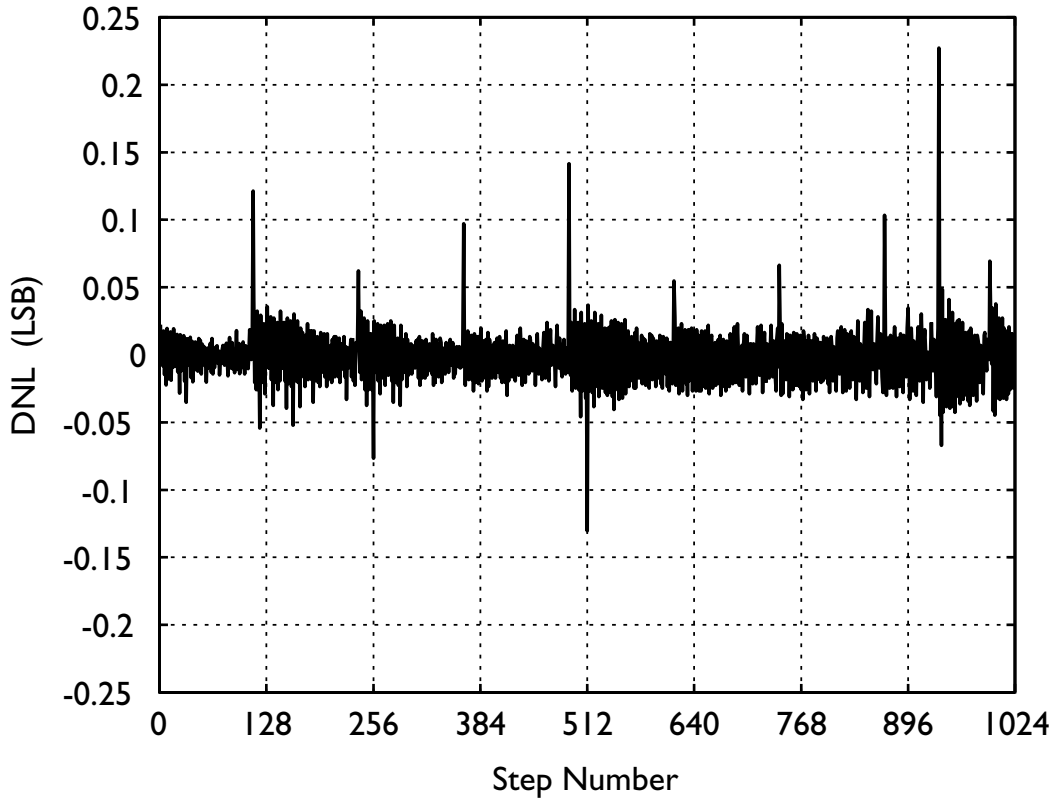

Figure 2.16: Measured static DNL of a 10 bit ladder

2.7 Summary

Linearity is one of the most important design issues of an A/D converter. Therefore, the main building block which determines the linearity of the entire converter, needs special attention. For many of the A/D converter architectures, e. g. successive-approximation converters or pipelined A/D converters, this building block is the D/A converter.

Modern deep submicron processes, in particular, require special additional processing steps to obtain high-precision passive elements. In this situation, it is advantageous to be able to build D/A converters which use MOS transistors exclusively as elements defining the accuracy.

The MOSFET-only ladder is a possible implementation of a D/A converter which uses MOS transistors only. To obtain maximum linearity, the second-order effects associated with this special type of ladder have to be understood and expressed analytically. These expressions can then be used as a basis for an optimized design.

The measurements performed on MOSFET-only ladders show that the matching capability of MOS devices is as good as the matching reported for precise passive components. Linearity of 11 bit and more is feasible. Therefore, MOSFET-only ladders can be used as a high-performance replacement of passive elements in nearly all converter implementations.

3

A 10 bit Successive-Approximation A/D Converter

Wie ist's? Will's fördern? Will's bald gehn?

FAUST, Der Tragödie erster Teil

Telecommunication has a strong need for A/D converters with a linearity of 12 bit to 14 bit, either for wireless or for traditional wire-bound communication systems [40]. The measurements of the MOSFET-only ladder presented in chapter 2 show that this range of resolution can be implemented using MOS transistors as ladder elements. As noted in chapter 1, although A/D converter implementations with precise passive elements such as resistors [41,42] or, more important in CMOS design, capacitors [7,43] are used extensively, MOSFET-only solutions are often preferable.

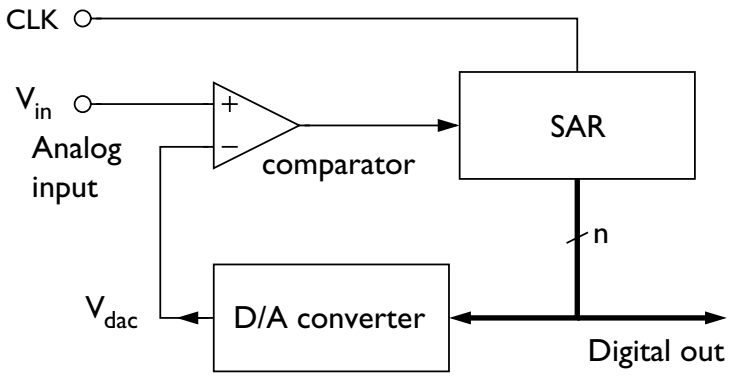

Figure 3.1: Block diagram of a successive-approximation A/D converter

As a final verification that the MOSFET-only ladder is capable of providing the measured performance not only statically, but also as the critical part in an A/D converter, a 10 bit successive-approximation A/D converter with a conversion speed of 200 ksample/s has been implemented. These parameters have been chosen to obtain a design where the focus lies on the maximum linearity achievable with the MOSFET-only ladder without trimming or calibration. The medium resolution of 10 bit relieves the overall design from stringent noise requirements, while an accuracy of the ladder beyond 10 bit can still be measured. The successive-approximation architecture, which has been selected for this design, is the best compromise in terms of accuracy, speed and silicon area for A/D converters in the medium resolution range and a conversion rate of several hundred ksample/s.

3.1 Successive-Approximation Architecture

A successive-approximation A/D converter basically consists of a feedback loop set up by a D/A converter, a comparator and the successive-approximation register (SAR), as shown in figure 3.1.

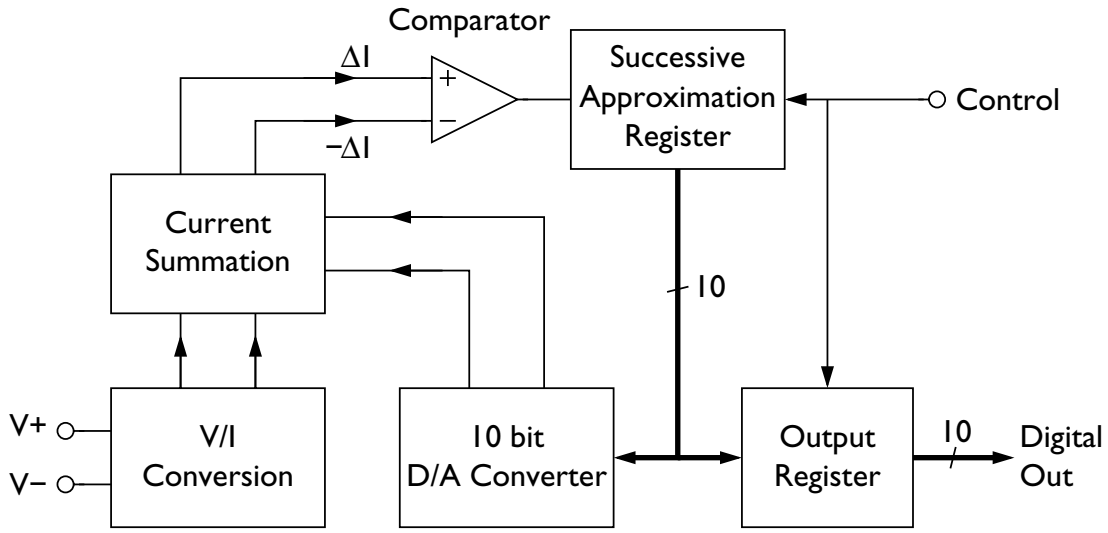

Figure 3.2: Detailed block diagram of the 10 bit A/D converter

The controlling element of the entire converter is the successive-approximation register. Prior to the start of a conversion cycle, the SAR is cleared. As a first step, the most significant bit (MSB) of the register is set, the other bits remain zero. This causes the D/A converter output to be in the middle of the conversion range of the A/D converter. If the comparator detects that the analog input voltage V_{in} is higher than the D/A converter output voltage, the MSB remains set, otherwise it is cleared. In the next clock cycle, the next significant bit is set, and the same decision is made. Successively, all subsequent bits are tested and eventually set. Thus, one conversion cycle of an n bit successive-approximation A/D converter needs at least n clock cycles before the output is determined. Normally, one additional clock cycle is consumed to clear the SAR before a new conversion cycle can start.

A more detailed block diagram of the implemented A/D converter is depicted in figure 3.2. The feedback loop is clearly visible. The entire signal path is kept fully differential. This reduces the interference between the analog and digital part of the circuit. Any spikes or substrate

noise stemming from the digital part enter the analog part as common-mode signals which are suppressed by the fully differential architecture. Moreover, differential signals double the effective signal amplitude and therefore increase the signal-to-noise ratio by 3 dB [44].

The following sections will describe the building blocks in detail, starting with those blocks that are most important for the overall performance of the A/D converter.

3.2 MOSFET-Only D/A Converter

The two building blocks that mainly determine the accuracy and dynamic behavior of the A/D converter are the D/A converter and the comparator. In today's technologies, however, the D/A converter is the bottleneck in achieving high accuracy. Therefore, the main focus of the design must be on this particular building block.

The D/A converter consists of the MOSFET-only ladder described in chapter 2. Based on the measurements presented in section 2.6, the W/L ratio of the transistors has been selected as 30 µm/30 µm. The reference current is 200 µA, which yields an LSB current of 195 nA for a 10 bit ladder. This value is a good compromise between power consumption on the one hand and an LSB value which can be handled easily by the subsequent circuit on the other hand.

Two questions arise when the MOSFET-only ladder is to be used in the context of an A/D converter. First, the signal which represents the analog information has to be chosen. The obvious choice is a voltage-based system, because the input signal of an A/D converter is almost always a voltage. The output signal of the ladder, on the other hand, is a current. Therefore, voltage-to-current or current-to-voltage conversion has to take place at one point of the circuit.

In this implementation, current has been chosen. The cost in terms of silicon area and power consumption is roughly the same for a current-to-voltage conversion of the ladder currents and the voltage-to-current

conversion of the analog input signals. Using currents has the advantage that the voltage swing of all internal nodes of the A/D converter is minimized and thus higher speed can be achieved for a given power consumption. Moreover, the voltage-to-current conversion at the analog input of the A/D converter can be accomplished by a single polysilicon resistor in contrast to the solution where the ladder output currents are converted to voltages, where precisely matched resistors would be necessary. The input voltage-to-current conversion circuit will be described in section 3.5.

The second question is the processing of the ladder current. The circuit extracting the ladder current has to fulfill two requirements. It was pointed out in section 2.3 that the ladder only functions properly if both terminals are at exactly the same voltage, otherwise an error in the output current occurs. This means that equation (2.27) represents an upper bound of the offset voltage V_{off} that can be tolerated in the extraction circuit.

The other requirement is the input impedance. The output current of the ladder is divided between the output impedance of the ladder and the input impedance of the node the current flows into. The error caused by this current division has to be below the LSB value of the output current. For a 10 bit ladder with a typical output impedance range of several kilohms to several tens of kilohms, this means that the input impedance has to be as small as a few Ohms.

The traditional way of collecting the current from an R-2R ladder and converting it back to an output voltage is the use of a transresistance based on an operational amplifier as illustrated in figure 3.3. This circuit is used extensively as an external current-to-voltage converter for integrated R-2R ladder-based D/A converters. It has the advantage that the output voltage range can be selected with a single external resistor.

This architecture, however, has two drawbacks. First, only half of the ladder current is used, the other half is simply dumped. This increases the power consumption of the circuit unnecessarily. Secondly, there is a conflict between achieving a low impedance at the operational amplifier

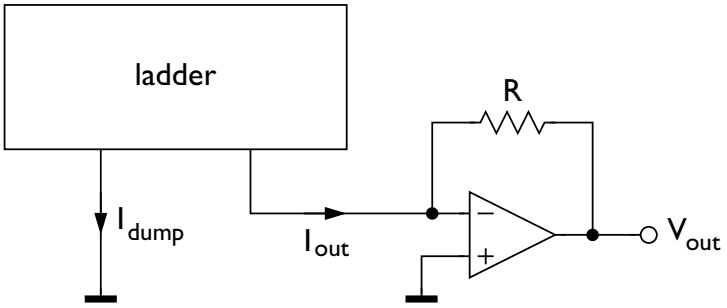

Figure 3.3: Traditional way of processing the ladder output current

input and a high transresistance. The transresistance, which equals R, should be made large enough so that the LSB current value generates a voltage that can be handled by a comparator without posing special requirements such as very low offset on that comparator. The input impedance, on the other hand, which is given by R divided by the gain of the operational amplifier, should be small for the reasons discussed above. For a typical LSB current value of 100 nA and a desired minimum LSB output voltage of about 10 mV, the transresistance has to be 100 kΩ. To achieve an input impedance of 1 Ω, on the other hand, the operational amplifier has to provide a gain of 100 dB, a value which is difficult (or at least costly) to achieve in practice, especially when driving a resistive load.

A solution to this conflict is to separate the functions of extracting the current from the ladder and converting the current to a voltage.

3.3 A Current Summation Circuit

The most important requirement for a circuit extracting the ladder current is a very low impedance at the input node, as discussed above. The circuit depicted in figure 3.4 provides this low input impedance. Transistors *T1*, *T3* and operational amplifier *OP1* form a regulated cascode structure [45, 46]. The input nodes where the ladder currents are

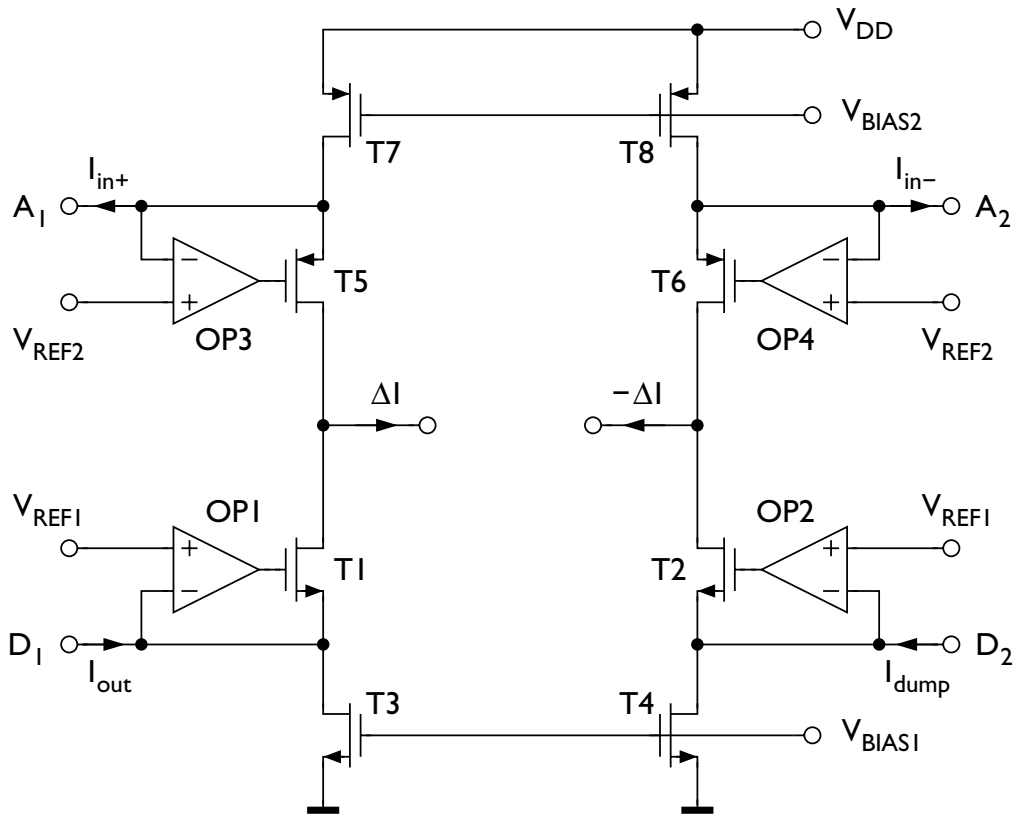

Figure 3.4: A current summation circuit to extract the ladder current accurately

collected are D_1 for I_{out} and D_2 for I_{dump}. The input impedance of these nodes is kept small by the loop of the regulated cascode and equals

$$r_{\text{in}} = \frac{1}{g_{\text{m1}}(A+1)} \tag{3.1}$$

where g_{m1} denotes the transconductance of transistor *T1*, and A is the gain of the operational amplifier. Equation (3.1) shows that this solution indeed allows to achieve a very low input impedance without being in conflict with any transresistance requirements.

The second requirement for proper current extraction, an offset low enough not to compromise the ladder accuracy, as discussed in sec-

tion 2.3, has to be addressed separately in the circuit of figure 3.4. The offsets of the two operational amplifiers *OP1* and *OP2* will cause the voltages at nodes D_1 and D_2 to differ slightly. Typical offset voltages for amplifiers in a CMOS technology are in the range of 5 mV to 10 mV. The limit presented in equation (2.27) of section 2.3 for the 10 bit ladder used in this implementation, however, yields a maximum offset voltage of 0.4 mV for a current error of 0.5 LSB. Therefore, some special circuitry has to be provided to reduce the offset.

One possibility to perform offset reduction is auto-zeroing. This solution has been implemented in the pipelined A/D converter presented in chapter 5. For the converter discussed here, the offset voltage of the amplifier *OP1* can be measured and adjusted externally. This solution has the advantage that the predicted influence of the offset voltage on the ladder can be verified experimentally. Measurements of the ladder performance with offset are presented in section 3.6.

3.3.1 Gain Requirements of the Amplifier

Although offset is the most important error which causes the voltages at nodes D_1 and D_2 to be different, even with offset-free operational amplifiers these voltages are not identical. Voltage differences occur due to the finite gain of the auxiliary amplifiers. A schematic of the regulated cascode including the most important voltages and currents is shown in figure 3.5. The actual voltage at node D_1 is not V_{REF1} as it should be ideally for an amplifier without offset, but $V_{\mathrm{REF1}} - \Delta V$. ΔV is the voltage at the input of the operational amplifier and equals its output voltage divided by its gain A. Because ΔV has the same effect on the accuracy of the ladder as an offset voltage of the amplifier would have, equation (2.27) can be used to calculate the maximum value ΔV_{max} that can be tolerated. With the aid of ΔV_{max}, the minimum gain A_{min} of the operational amplifier which is necessary to maintain the ladder accuracy can be calculated. Using the notation in figure 3.5, A_{min} can

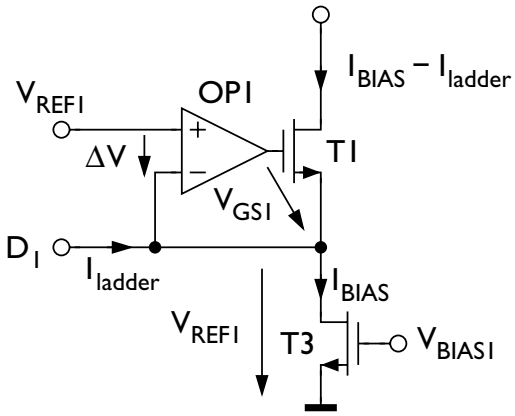

Figure 3.5: Amplifier gain calculation for the regulated cascode

be written as

$$A_{\min} = \frac{V_{\mathrm{GS1}} + V_{\mathrm{REF1}}}{\Delta V_{\max}} \tag{3.2}$$

Substituting V_{GS1} leads to the following equation:

$$A_{\min} = \frac{V_{\mathrm{T1}} + \sqrt{\dfrac{2I_{\mathrm{BIAS}}}{\mu_{\mathrm{eff}} C'_{\mathrm{ox}} \frac{W_1}{L_1}}} + V_{\mathrm{REF1}}}{\Delta V_{\max}} \tag{3.3}$$

V_{T1} denotes the threshold voltage of *T1* including the bulk effect, and I_{BIAS} is the bias current defined by *T3* which is the maximum value of the current that flows through *T1*. Equation (3.3) is a worst-case estimation of the required gain. The reason is as follows. Any value of ΔV occurring at both sides of the circuit, that is at the inputs of *OP1* and *OP2* in figure 3.4, respectively, only causes a shift of the reference voltage, but does not introduce any errors because both terminals of the ladder still are at the same voltage. ΔV can be written as

$$\Delta V = \frac{V_{\mathrm{T1}} + V_{\mathrm{DSsat1}} + V_{\mathrm{REF1}}}{A} \tag{3.4}$$

where the output voltage of the operational amplifier has been split into three parts. Assuming identical gain of the amplifiers *OP1* and *OP2* and identical transistors *T1* and *T2*, the portion of ΔV caused by V_{T1} and V_{REF1} would be identical on both sides and thus would not cause any errors. This argumentation, however, only holds if both amplifiers have identical gain, or at least if the gain mismatch can be estimated. Unfortunately, the gain of an operational amplifier is one of the least controlled parameters, and data on the gain matching is usually not available. Nevertheless, equation (3.3) can be used as a worst-case estimate assuming that the amplifier of the other side has infinite gain.

3.3.2 Speed Requirements

The second parameter of the operational amplifier which has to be examined, after having determined the minimum required gain, is the unity-gain bandwidth. For this purpose, the dynamic behavior of the input impedance of the regulated cascode has to be considered. The step response of the regulated cascode is calculated, using a small-signal equivalent circuit of the schematic shown in figure 3.5. The following assumptions are made. The output current difference ΔI in figure 3.4 will be fed into a current comparator with low input impedance. Thus, the drain of *T1* in figure 3.5 can be connected to analog ground in the small-signal equivalent circuit. The amplifier is assumed to be a single stage operational transconductance amplifier (OTA). The behavior of the voltage at node D_1 has to be examined when a current step is applied to that node. The small-signal equivalent circuit resulting from these assumptions is shown in figure 3.6. I_{in} is the input current step, r_1, g_{m1} and C_{gs1} are the small-signal output resistance, transconductance and gate-source capacitance of *T1*, respectively. r_3 denotes the output resistance of *T3*, and C_3 is the sum of all capacitances at the drain of *T3*. The single-stage OTA is described by its transconductance G_{m} and the dominant pole modeled with the output resistance R_{o} and the load capacitance C_{L}. The transfer function is found to have a single zero in

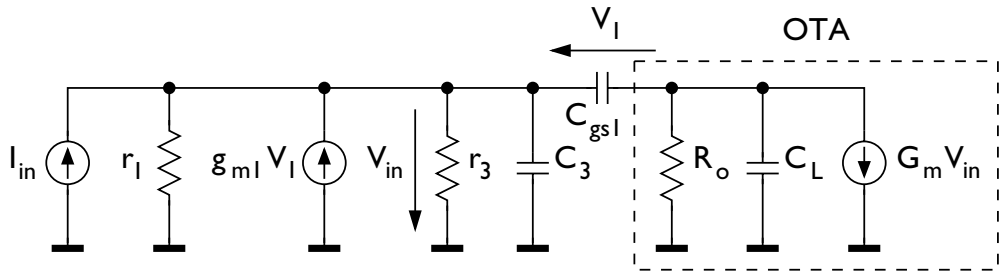

Figure 3.6: Small-signal equivalent circuit of the regulated cascode

the left half-plane and two poles:

$$V_{\text{in}} = \frac{I_{\text{in}}}{A\, g_{\text{m1}}} \frac{1 + s\dfrac{A}{\omega_{\text{GBW}}}}{\left(1 + \dfrac{s}{\omega_{\text{GBW}}}\right)\left(1 + \dfrac{s}{\omega_2}\right)} \tag{3.5}$$

with the following definitions:

$$A = R_{\text{o}} G_{\text{m}} \tag{3.6}$$

$$\omega_{\text{GBW}} = \frac{G_{\text{m}}}{C_{\text{L}}} \tag{3.7}$$

$$\omega_2 = \frac{g_{\text{m1}}}{C_3 + C_{\text{gs1}}} \tag{3.8}$$

where A and ω_{GBW} denote the gain and gain-bandwidth product[1] of

1 In this text the *gain-bandwidth product* is used in all calculations because it simplifies the mathematics. The gain-bandwidth product is defined as the DC gain of the amplifier multiplied by the dominant pole frequency. More important for the design of an amplifier, however, is the *unity-gain bandwidth*, which is the frequency at which the gain drops to unity. But if the phase margin of the operational amplifier is sufficiently large, the difference between the two frequencies is small. This is the reason why no distinction is made between the two definitions in most books [21, 47].

the OTA, respectively. The step response can be calculated from equation (3.5) and is given by

$$V_{\text{in}} = \frac{I_{\text{in}}}{A\, g_{\text{m1}}} \cdot \left[1 + \frac{A-1}{1-\dfrac{\omega_{\text{GBW}}}{\omega_2}}\, \mathrm{e}^{-t\omega_{\text{GBW}}} - \frac{A-\dfrac{\omega_{\text{GBW}}}{\omega_2}}{1-\dfrac{\omega_{\text{GBW}}}{\omega_2}}\, \mathrm{e}^{-t\omega_2} \right] \tag{3.9}$$

In this design, which is implemented in a 1 µm technology, the non-dominant pole ω_2 is much larger than ω_{GBW}. This is due to the fact that g_{m1} has been chosen large to reduce the factor before the brackets in equation (3.9) and therefore minimize the overall error. In conjunction with small capacitances C_3 and C_{gs1}, this results in a non-dominant pole at high frequencies. Therefore, the settling of the last term in equation (3.9) is very fast and can be neglected.

The gain-bandwidth product ω_{GBW} of the auxiliary amplifier has to be chosen such that the input voltage of the regulated cascode has settled at the end of the clock cycle. This is equivalent to demanding that the static term in equation (3.9), which is the DC input impedance of the regulated cascode, dominates the error at the end of the clock cycle. For auxiliary amplifiers with a typical gain of 60 dB to 80 dB this requirement is fulfilled if the gain-bandwidth product ω_{GBW} is chosen two to three times the internal clock frequency $2\pi f_{\text{CLK}}$.

3.3.3 Output Current Generation

In a successive-approximation architecture, as discussed in section 3.2, the comparator has to decide whether the analog input signal or the D/A converter output is higher (see figure 3.1). In other words, it decides whether the difference between these two signals is positive or negative. In a fully differential system, as is the case in this implementa-

tion, the difference between two differential signals has to be evaluated. This can be achieved by either using a fully differential comparator with two differential input pairs, or the subtraction of the two differential signals can be accomplished before feeding the difference to the comparator.

The second solution can be implemented in this A/D converter in an elegant way because currents are used as signals and therefore subtraction can be accomplished easily. The signal path is shown in the block diagram of the converter (figure 3.2). To generate the current difference between the ladder output current and the analog input current, denoted ΔI and $-\Delta I$ in figure 3.4, the regulated cascode structure is mirrored to the V_{DD} side. The regulated nodes of these cascodes (denoted A_1 and A_2 in figure 3.4) are used to feed in the analog input current, which directly comes from the input voltage-to-current converter and is also differential.

3.4 A Current-Input Comparator

3.4.1 Current Comparator Structures

The input of the comparator in the successive-approximation loop is a fully differential current which equals the difference between the input signal and the D/A converter signal, as described in the last section. Typical values of the LSB current difference are of the order of 100 nA. The comparator must be sufficiently sensitive to handle such small currents at the required speed.

A current comparator would be the natural solution. In contrast to a voltage comparator, which essentially is an amplifier with high gain, the current comparator is a transresistance amplifier, because it has to provide an output voltage to be compatible with the digital circuitry that senses the output signal. Therefore, an ideal current comparator would have to provide infinitely small input impedance and infinite transresis-

tance, two requirements which are in conflict, as already mentioned in section 3.2.

Several circuits have been published which utilize different solutions to transform the input current to an output voltage, such as current mirrors and the output resistance of MOS transistors [48, 49], resistors [50] or a CMOS cross-coupled latch [51]. The main problem, however, remains the implementation of the large transresistance that is required.

3.4.2 Current-to-Voltage Conversion at the Input

Another solution is to move the current-to-voltage conversion in front of the comparator where the requirements for the magnitude of the transresistance are less stringent. The input current difference is converted to a voltage, which is then fed to a conventional voltage comparator to make the logic decision. The advantage of this approach is that large gain and high speed can be achieved simultaneously. Moreover, well-known architectures can be chosen for the voltage comparator part.

Since the comparator provides a binary output, linearity in the voltage-to-current conversion is not required. Non-linearity is even desired: A high transresistance can be chosen for small currents to produce a large LSB voltage at the input of the comparator, and the transresistance can be decreased gradually towards higher currents to clamp the voltage swing of the input node in order to increase the speed. If the comparator is fed with a fully differential current difference, as is the case in this implementation, not even matching between the transresistance elements of the two converter inputs is necessary because both input currents will always exhibit opposite polarity.

A circuit to implement this non-linear current-to-voltage conversion is shown in figure 3.7. Transistor *T3* is the actual transresistance device, operating in linear region. V_{CM} is the common-mode voltage of the comparator inputs. The value of the equivalent resistance of *T3* is chosen for maximum LSB voltage. For this design, the LSB current of

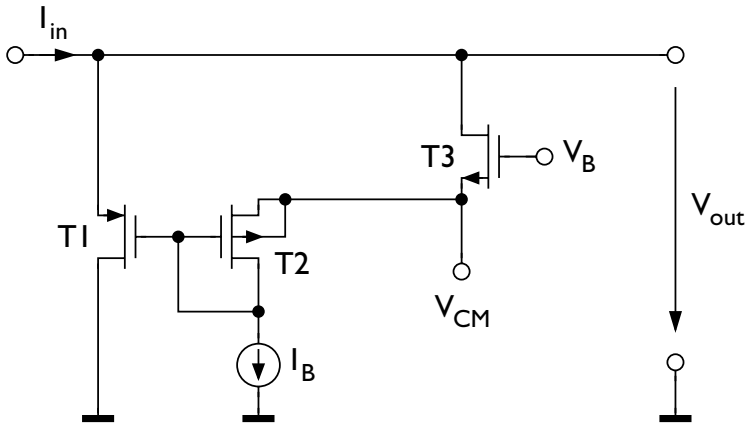

Figure 3.7: Current-to-voltage conversion at the comparator input

195 nA produces a comparator input voltage of 10 mV. For large positive currents I_{in} (in direction of the arrow in figure 3.7) transistor *T3* would be forced out of the linear region, resulting in a large voltage at the comparator input. To prevent this situation, a clamping network consisting of *T1* and *T2* starts to conduct current and clamps the output voltage. Since transistor *T1* has a source-bulk voltage larger than zero (in contrast to *T2*), the threshold voltage of *T1* is always higher than that of *T2* due to the bulk effect. This guarantees that the clamping always takes place at a voltage above V_{CM}, independent of process variations. *T1* is the actual clamping device, whereas *T2* only is used to bias *T1* properly. For negative input currents no such clamping is necessary: A higher input current enlarges the gate-source voltage of the transresistance element *T3* and therefore lowers its resistance.

3.4.3 The Voltage Comparator

The voltage comparator is depicted in figure 3.8. Transistors *T1* and *T2* form a differential pair which acts as a preamplifier. The devices *T3*, *T4*, *T5* and *T6* form a cross-coupled CMOS latch with outputs Q and $\overline{Q}$. These two outputs feed an RS-flipflop to obtain stable outputs

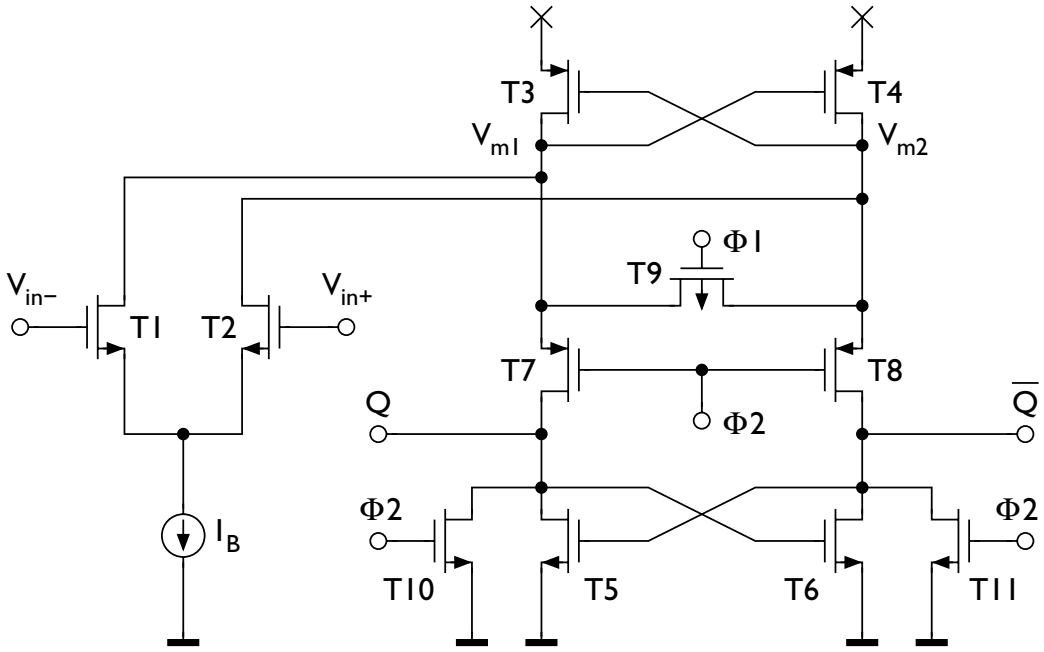

Figure 3.8: The core voltage comparator

during the entire clock phase, even when the comparator is reset. A two-phase clocking scheme is used to reduce comparator offset [52]. The two clock phases are depicted in figure 3.9.

Regeneration of the PMOS cross-coupled latch starts with the rising edge of signal Φ1 which opens switch *T9* (see time ❶ in figure 3.9). It can be seen in figure 3.9 that at this time the second clock Φ2 is still high. This keeps the NMOS cross-coupled latch (*T5* and *T6*) in the reset mode and disconnects it from the PMOS latch via the two switches *T7* and *T8*. After the PMOS latch has established some voltage gain, the NMOS latch is connected to the PMOS side when Φ2 goes low (time ❷) to generate a full-swing output. The advantage of this clocking scheme is as follows. During the first short regeneration phase where Φ1 and Φ2 are high only the PMOS latch is active and contributes to the offset. When the NMOS latch is connected in the second phase, gain has been established already, which reduces the offset contribution

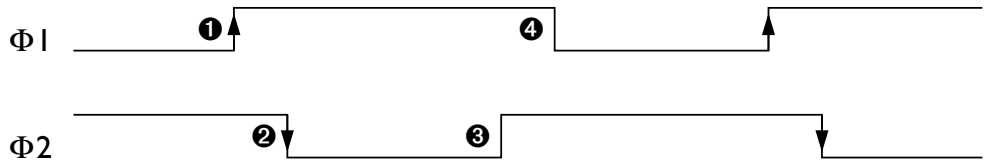

Figure 3.9: Clocking scheme of the comparator: ❶ start of the regeneration of the PMOS latch, ❷ full regeneration of the entire latch, ❸ outputs become inactive, ❹ reset of the comparator

of transistors *T5* to *T8* and *T10* and *T11*. Therefore the input-referred offset is mainly caused by mismatch between transistors *T1*, *T2* and *T3*, *T4*.

The offset caused by transistors *T3* and *T4* depends on the absolute value of the gain from the input to the PMOS cross-coupled latch, which equals

$$|A| = \frac{g_{m1}}{g_{m3}} \tag{3.10}$$

at the moment the regeneration starts. To reduce the offset contribution of transistors *T3* and *T4*, the gain must be larger than unity. In this design, a gain of five has been realized.

To obtain high speed, it is important to optimize the time constant τ_{reg} of the first regeneration step. The calculation is straightforward and yields

$$\tau_{reg} = \frac{C_m}{g_{m3}} \tag{3.11}$$

where C_m denotes the sum of all capacitances at node V_{m1} and V_{m2} in figure 3.8, respectively. The dimensions of transistor *T3* not only determine g_{m3} but also affect C_m. Therefore, τ_{reg} can be optimized by optimizing the width W_3 of *T3* (assuming fixed transistor lengths). Capacitance C_m is composed of contributions of *T1*, *T3*, *T7* and *T9*. Because the dimensions of *T1* and *T3* are linked via equation (3.10), the

contribution of *T1* can be expressed as a function of W_3. Therefore, C_m can be written as

$$C_m = \zeta W_3 + C_s \tag{3.12}$$

where ζ and C_s depend on process parameters and the dimensions of the other transistors. C_s is the part of C_m that is independent of W_3. The optimum width W_3 can now be calculated and yields

$$W_{3_{opt}} = \frac{C_s}{\zeta} \tag{3.13}$$

which means that the optimum is reached if the part of C_m depending on W_3 is equal to the independent portion of C_m.

The current-input comparator which results from combining the input current-to-voltage conversion and the subsequent voltage comparator is characterized by high speed and low offset. The speed is achieved by clamping the input voltage swing in the current-to-voltage conversion circuit and by optimizing the design of the regenerative comparator. The low offset is a result of three measures. The large transresistance of the current-to-voltage conversion causes a large LSB voltage at the input of the voltage comparator. The two-phase clocking scheme reduces the number of devices that contribute to the offset, and finally the choice of appropriate g_m-ratios further reduces the input-referred offset. The result is a fast, reliable and small current-input comparator.

3.5 Input Voltage-to-Current Conversion

The last building block of the A/D converter to be described is the voltage-to-current conversion at the analog input. This conversion is necessary because current has been chosen as signal throughout the converter, as has been discussed in section 3.2.

The circuit is shown in figure 3.10. To obtain maximum linearity, a polysilicon resistor has been chosen as the transresistance element. The

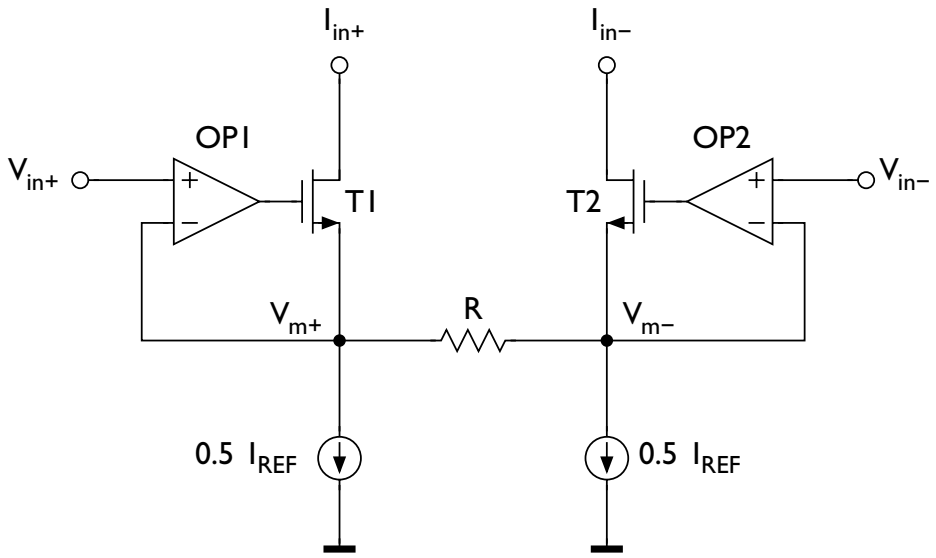

Figure 3.10: Input voltage-to-current conversion

reference current of 200 µA and the full scale input voltage of 2 V yield a resistance of 10 kΩ. This value can be implemented comfortably even if the only resistive layer available is relatively low-ohmic polysilicon (the sheet resistance $R_\square$ of polysilicon is 33 Ω/□ in the technology used for this implementation). The absolute value of this resistor is not critical because it only changes the conversion range of the A/D converter, but does not affect linearity. Thus, even silicided polysilicon can be used for this purpose. This type of polysilicon is used in modern deep submicron processes and has a higher variance of the sheet resistance than unsilicided polysilicon [20].

The differential input voltages $V_{\text{in}+}$ and $V_{\text{in}-}$ are converted to differential input currents $I_{\text{in}+}$ and $I_{\text{in}-}$. The value of the two current sources in figure 3.10 equals half of the full scale reference current I_{REF}. Differential currents are generated even if one of the inputs is held at a fixed reference voltage, while the input signal is applied to the other terminal. This means that the circuit works as a single-ended to differential converter in this case which is convenient for measurement

setups. The transistors, operational amplifiers and the reference current sources form two regulated cascodes similar to those described in section 3.3.

To obtain the full accuracy, the operational amplifiers have to provide a certain minimum gain. Although the discussion in general follows the calculation carried out for the current summing circuit in section 3.3.1, some important differences exist.

Due to the finite gain A of the operational amplifiers the voltage across resistor R in figure 3.10 is not exactly the difference between the input voltages $V_{\text{in}+}$ and $V_{\text{in}-}$. The reason is the small voltage ΔV at the input of the operational amplifiers, as discussed in section 3.3.1.

In contrast to the calculation for the current summing circuit, the changing input voltage which also changes the voltage at nodes $V_{\text{m}+}$ and $V_{\text{m}-}$ in figure 3.10 has to be taken into account in this case. The maximum voltage error $\Delta V_{\max}$ that can be tolerated without degrading the performance of the A/D converter is 0.5 LSB of the input voltage. The maximum value of ΔV on one side of the voltage-to-current conversion circuit occurs if the input voltage at that side is also at its maximum. To obtain a worst-case estimation, it is assumed that the operational amplifier on the other side has infinite gain such that no error occurs at that side (the same argument has been used in section 3.3.1). Therefore, it is sufficient if the voltage error ΔV on one side of the input circuit does not exceed 0.5 LSB. Any error on the other side will only decrease the total voltage error. The following calculation is carried out for the positive input terminal, the results for the negative terminal are identical.

Similar to equation (3.2), the minimum gain can be written as

$$A_{\min} = \frac{V_{\text{GS1}} + V_{\text{in}+_{\max}}}{\Delta V_{\max}} \tag{3.14}$$

where $V_{\text{in}+_{\max}}$ denotes the maximum voltage at the positive input terminal. The substitution of V_{GS1} in this case has to take into account the bulk effect because the source of transistor *T1* (which is at node $V_{\text{m}+}$

in figure 3.10) is exposed to the full swing of the input voltage. The following equation is used to model the bulk effect [28]:

$$V_{\mathrm{T}} = V_{\mathrm{T0}} + \gamma\left(\sqrt{2\phi_{\mathrm{F}} + V_{\mathrm{SB}}} - \sqrt{2\phi_{\mathrm{F}}}\right) \tag{3.15}$$

where V_{T0} denotes the zero-bias threshold voltage, ϕ_{F} is the Fermi potential of the doped silicon bulk and V_{SB} denotes the source-bulk voltage. Although above equation is not very accurate, it is sufficient in this case. Using equation (3.15) in (3.14) to substitute V_{GS1} and noting that the current through *T1* is I_{REF} for the maximum input voltage, the minimum required gain can be written as

$$A_{\mathrm{min}} = \frac{1}{\Delta V_{\mathrm{max}}}\left[V_{\mathrm{in+_{max}}} + \sqrt{\frac{2I_{\mathrm{REF}}}{\mu_{\mathrm{eff}} C'_{\mathrm{ox}} \frac{W_1}{L_1}}} + V_{\mathrm{T0}} + \gamma\left(\sqrt{2\phi_{\mathrm{F}} + V_{\mathrm{in+_{max}}}} - \sqrt{2\phi_{\mathrm{F}}}\right)\right] \tag{3.16}$$

The required gain-bandwidth product ω_{GBW} of the operational amplifiers can be calculated using the equations given in section 3.3.2.

A second building block which can affect the accuracy of the input voltage-to-current conversion are the two current sources in figure 3.10. The voltage across these current sources changes with the input voltage of the A/D converter, therefore their output resistance must be high enough such that the accuracy of the current is not compromised. If the input circuit is used as a single-ended to differential converter, things are even worse because one of the current sources sees the full differential input voltage swing. The current, however, must not change by more than half of an LSB, which poses a stringent lower limit on the output impedance of the current source. For this A/D converter with a full scale input voltage of 2 V and an LSB current of 195 nA, the output resistance has to be more than 20 MΩ.

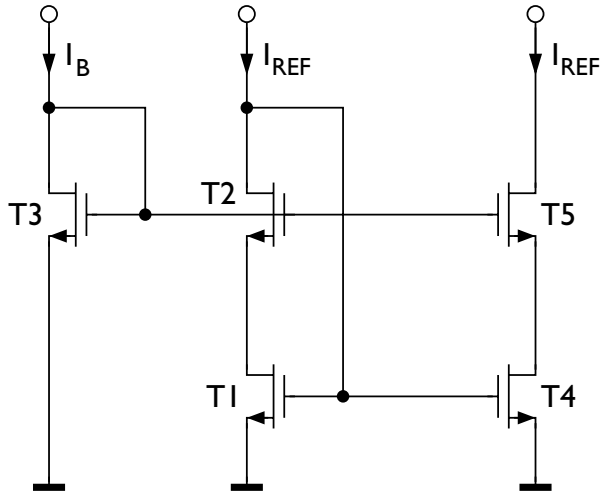

Figure 3.11: High-swing cascode for the input current sources

An output impedance of this order of magnitude cannot be implemented with a single transistor. A conventional cascode, however, is also infeasible because the voltage drop that is allowed over the current sources in figure 3.10 is too low for the conventional cascode biasing. Therefore, a high-swing cascode has been used in this implementation, as depicted in figure 3.11 [53]. Using a channel length of 5 µm, the current source shows an output resistance of 40 MΩ in simulation at a minimum output voltage of 450 mV.

3.6 Implementation and Measurements

The A/D converter described in this chapter has been implemented in a standard digital 1 µm CMOS process with a single layer of polysilicon and two metal layers [37, 54]. The chip micrograph is depicted in figure 3.12. The MOSFET-only ladder is clearly visible in the middle of the core. The chip area including the pad ring is 5.15 mm^2, the core area measures 2.15 mm^2.

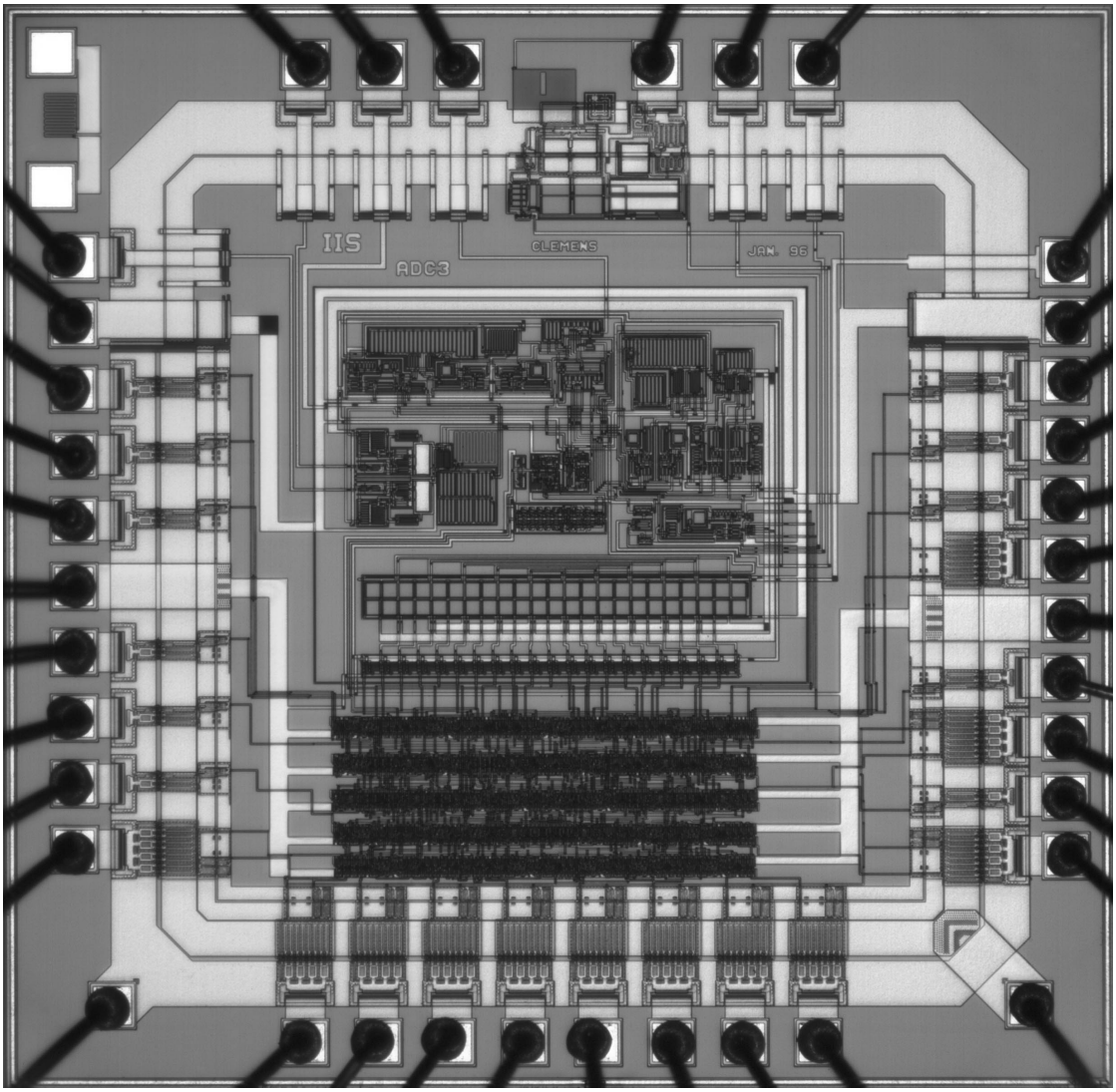

Figure 3.12: Chip micrograph of the 10 bit A/D converter

It has been one of the goals of this implementation to investigate whether MOSFET-only ladders are capable of delivering the same accuracy as converters using passive elements. One possibility of comparison is the total harmonic distortion (THD) that can be achieved with these converters. For a 10 bit A/D converter, the THD should be below −60 dB [55].

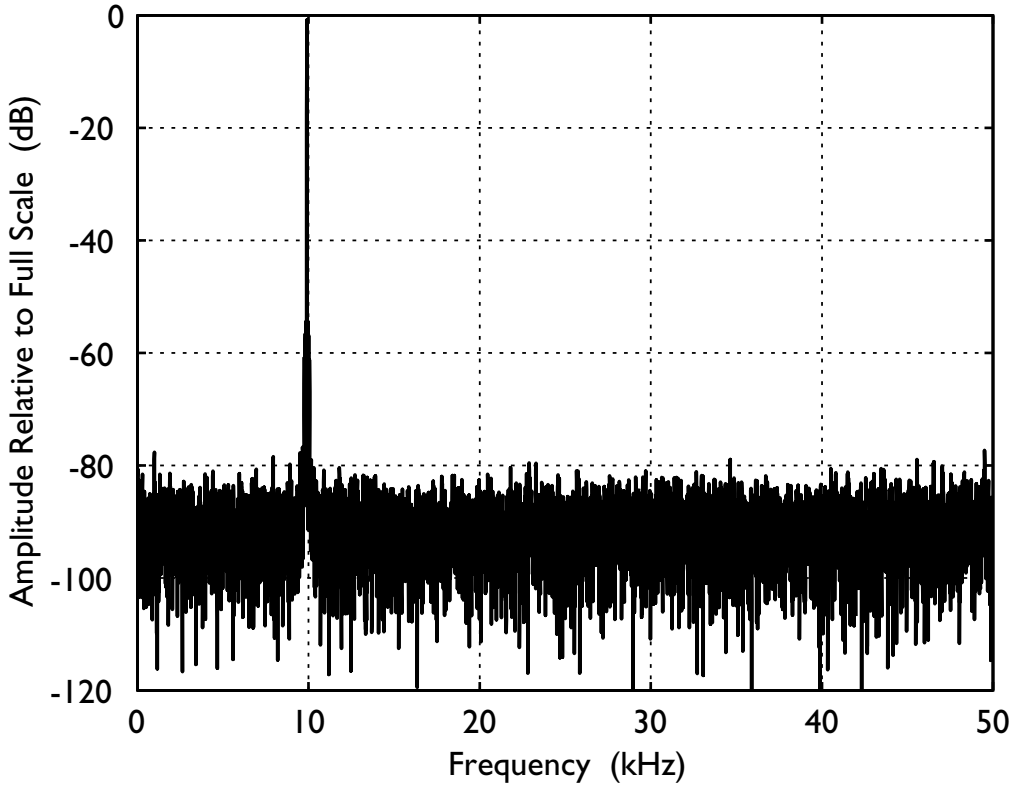

Figure 3.13: Measured output spectrum of the A/D converter. The input frequency is 9.997 kHz, sampled at 100 ksample/s using a 16384 point FFT with a Kaiser-Bessel window (α = 4.0).

Figure 3.13 shows the output spectrum of a full scale sine wave of 9.997 kHz, sampled at 100 ksample/s. The THD, which has been measured after averaging, is −79 dB. Table 3.1 shows the measured THD of nine samples. This data shows that the measured THD of all samples exceeds the values expected for a 10 bit A/D converter.

The maximum speed of the A/D converter is 200 ksample/s, which translates to an internal clock frequency of 2.6 MHz. Although the successive-approximation register of the converter is capable of performing the setup at the beginning of a conversion without additional clock cy-

Table 3.1: THD measurements of different samples

Measured THD	Number of Samples
−79 dB	1
−76 dB	2
−70 dB	3
−68 dB	2
−67 dB	1

Table 3.2: Effective number of bits (ENOB) for different sampling frequencies

Sampling frequency	Measured ENOB
10 kHz	9.8 bit
100 kHz	9.4 bit
200 kHz	9.1 bit

cles, three additional clock cycles have been used in the measurement setup for the sample-and-hold operation and input settling.

A second measure for the dynamic performance of an A/D converter is the effective number of bits (ENOB). This is the number of bits an ideal A/D converter would have if its signal-to-noise ratio is identical to the signal-to-noise ratio of the converter under test. The measured effective number of bits for different sampling frequencies are summarized in table 3.2. The input frequencies of these measurements were chosen to be one tenth of the respective sampling frequency.

The offset sensitivity of the MOSFET-only ladder has been investigated in section 2.3. To measure the effects of offset on the ladder accuracy, one operational amplifier of the current summing circuit has an externally adjustable offset voltage to be able to apply arbitrary offset voltages to the ladder (see section 3.3). Figure 3.14 shows the spectrum

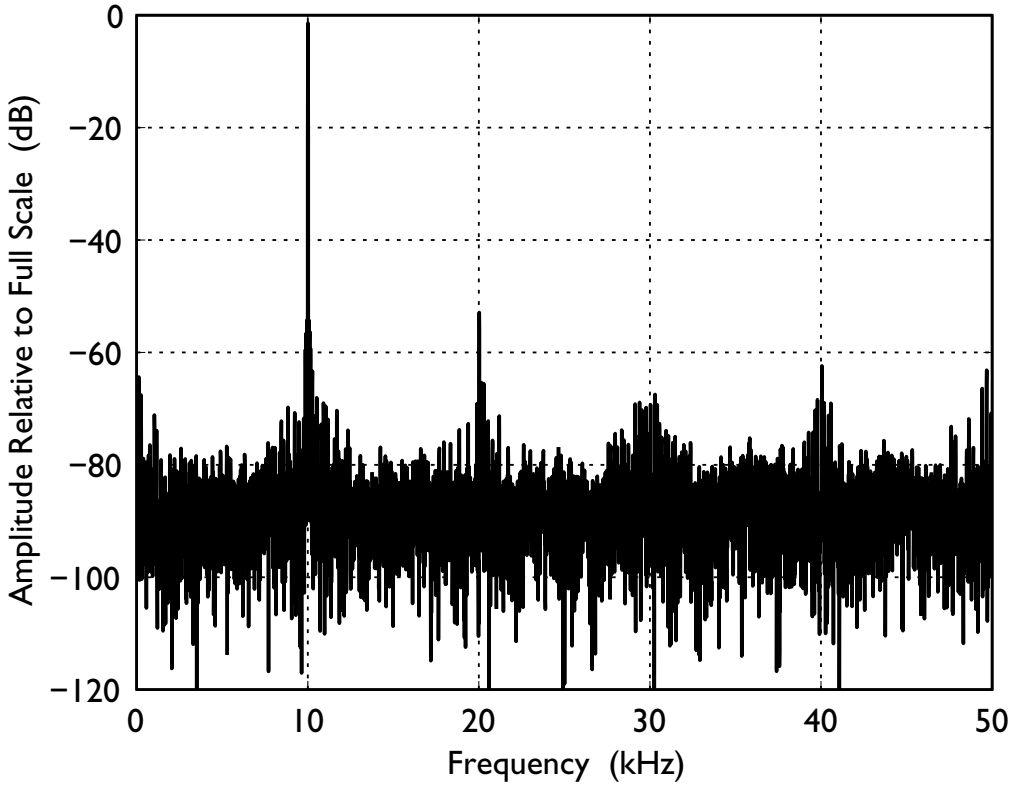

Figure 3.14: A/D converter output at 100 ksample/s with 10 mV offset applied to the ladder output

of the A/D converter with the amplifier offset adjusted to 10 mV. This amount of offset is a typical value for CMOS amplifiers which do not feature any offset cancellation scheme. The measured THD is −52 dB which corresponds to 8 to 9 bit linearity, which is in close agreement with the prediction of equation (2.27) in section 2.3.

The A/D converter operates from a single 5 V power supply to stay compatible with standard digital circuitry. The power consumption, including the analog and digital part as well as a bandgap reference and the digital output buffers is 12 mW. A summary of the A/D converter characteristics is given in table 3.3.

Table 3.3: A/D converter characteristics

Process	1 µm Single-Poly CMOS
Resolution	10 bit
Input Voltage Range	2 V peak-peak
Conversion Speed	200 ksample/s
Core Area	1490 µm × 1446 µm
Chip Area	2281 µm × 2257 µm
Power Supply	Single 5 V
Power Consumption at 200 ksample/s	12 mW with 5 V supply

3.7 Summary

The A/D converter described in this chapter has achieved its goal in showing that the MOSFET-only ladder is capable of providing high linearity not only in static measurements of the ladder alone, but also as critical dynamic part of an A/D converter. An untrimmed THD as low as −79 dB has been measured. This proves that MOS transistors are capable of precise matching even in large arrays, provided that special care is taken in circuit development and layout. The measured matching accuracy of the transistors is as good as comparable circuits using capacitor or resistor arrays.

A successive-approximation architecture has been chosen, which is the best compromise in terms of speed, complexity and silicon area for A/D converters with medium resolution and speeds up to several hundred ksample/s.

Although the D/A converter is the most important building block in a successive-approximation architecture that determines the final linearity of the entire A/D converter, it is mandatory that the remaining building blocks of the converter are designed with the same care as the D/A converter. Otherwise, the dynamic performance of the A/D converter will not be able to match the static performance of the D/A converter.

Having proved the high performance of the MOSFET-only ladder in a medium-speed A/D converter design, the second design challenge in A/D converter design, namely achieving high speed, can be set about. A significant increase in speed is only possible with a different A/D converter architecture, which will be described in the next chapters.

4

Pipelined A/D Converter Architecture

Das heiß' ich endlich vorgeschritten!

MEPHISTOPHELES, Der Tragödie zweiter Teil

The A/D converter presented in chapter 3 demonstrates that the MOSFET-only ladder is capable of providing excellent matching accuracy. It has been stated in the introduction of chapter 3 that the main goal of the successive-approximation A/D converter design was to demonstrate that high linearity can be achieved using the MOSFET-only ladder. The conversion speed of this converter, although appropriate for many applications, is too low for modern telecommunications systems such as new wireline services like ADSL (Asymmetric Digital Subscriber Line) [56, 57]. The next step in the development described in this thesis is therefore the increase of conversion speed. Again, the focus will mainly be on the performance of the MOSFET-only ladder.

4.1 Increasing Conversion Speed

4.1.1 Enhanced Successive-Approximation Architectures

The natural way of increasing the speed of a successive-approximation A/D converter is to use a more sophisticated algorithm to perform the successive approximation. Although the basic algorithm implements the optimum search strategy for this kind of problem, an A/D converter with n bit resolution needs at least n clock cycles to complete one conversion. To overcome this limitation but nevertheless conserve the simple and straightforward successive-approximation architecture as much as possible, several enhancements to the basic algorithm have been proposed. Two different strategies are possible to increase the conversion speed: either to decrease the time needed per clock cycle or to maintain the clock frequency and decrease the number of clock cycles per conversion. The first solution is achieved by taking the decision of the comparator before settling to full accuracy has taken place and correcting the deliberately introduced error afterwards. This can be done either digitally [58] or in the analog domain [59, 60].

The second approach, i. e. decreasing the number of clock cycles needed, leads to a different solution. If only a reduced number of clock cycles is to be used for conversion, more than one bit has to be decided upon during each clock cycle. Although in principle an arbitrary number of bits can be evaluated simultaneously, in practice the limit will be two bits. If more than two bits were evaluated per clock cycle, the overhead in circuitry and silicon area would be too high to justify the successive-approximation topology. Although several architectures utilizing this approach have been proposed [61, 62], no implementation of such an enhanced A/D converter has been published to date.

The major problem of all enhanced successive-approximation algorithms is the rather limited increase in speed. A factor of four decrease in conversion time has been reported in [59], and a doubling of the conversion speed is estimated for the architectures that determine two bits per clock cycle [61, 62]. To substantially increase the conversion speed, a different solution has to be chosen.

4.1.2 Parallel A/D Converters

An alternative way of achieving high sampling rates is the use of a parallel architecture. With this approach, several identical A/D converters operate in parallel in a time-interleaved manner, as depicted in figure 4.1. The first converter processes the first input sample, the second converter the next one and so on until, after the last converter has processed its respective sample, the first converter has its turn again. The individual A/D converters therefore operate on a much lower sampling rate than the entire converter. The reduction in conversion speed for each individual converter is equal to the number of A/D converters in parallel. The only building block that sees the full input signal bandwidth of the composite converter is the sample-and-hold circuit of each A/D converter.

The parallel architecture can be used to increase the sampling speed of all common A/D converter architectures such as successive-approximation types or pipelined converters [13, 63, 64]. But apart from the advantage that existing designs can be used to construct a fast A/D converter, the parallel architecture has also some important drawbacks.

Increasing the conversion speed by doubling circuitry instead of increasing the speed of the individual building blocks leads to high power consumption and a large silicon area requirement. One possibility to keep the power consumption within tolerable limits is to share certain building blocks, especially operational amplifiers, between the individual A/D converters [64].

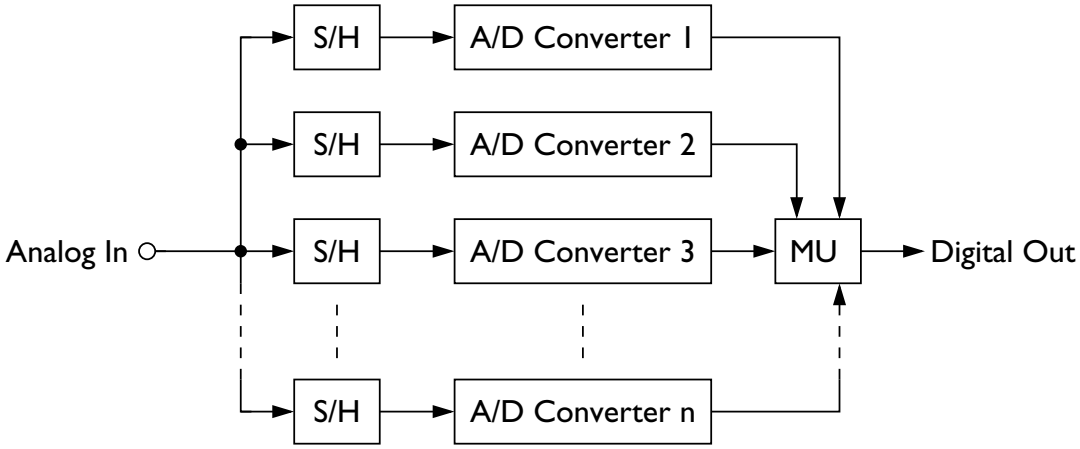

Figure 4.1: A parallel architecture for high-speed A/D converters

A second problem associated with parallel A/D converters is path mismatch. During operation, the input signal has to pass different paths from the input to the digital output. If all A/D converters in parallel are identical, these paths are also identical. But if offset and gain mismatch occur between the individual converters, the path for the input signal changes each time it is switched from one converter to another. This behavior gives rise to fixed-pattern noise at the output of the composite A/D converter which can be detected as spurious harmonics in the frequency domain [65].

The parallel architecture is advantageous when high sampling rates are necessary which are difficult to achieve with single A/D converters. Although the architecture is straightforward, parallel A/D converters usually are not the best compromise when it comes to increasing the conversion rate of medium speed converters. For the A/D converter family described in this thesis, it has therefore been decided to abandon the successive-approximation architecture in favor of a pipelined converter to obtain higher speed.

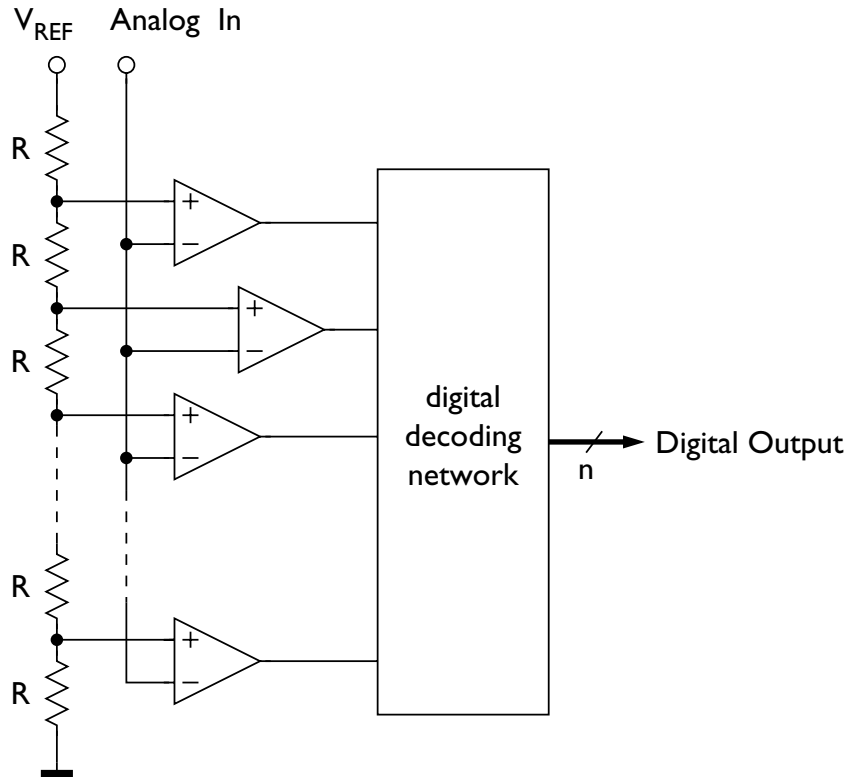

Figure 4.2: Basic architecture of a flash A/D converter

4.2 Pipelined A/D Converters

4.2.1 Flash Converters

In the last couple of decades, several A/D converter architectures have been developed in order to achieve high conversion speed. The first dedicated high-speed A/D converter was the flash converter which is still the fastest architecture available [66, 67].

In an n bit flash converter, the analog input voltage is compared to all $(2^n - 1)$ reference voltage levels simultaneously, as depicted in figure 4.2. The $(2^n - 1)$ reference voltage levels are generated by means of a resistor string. While this parallel comparison gives rise to high conversion rates,

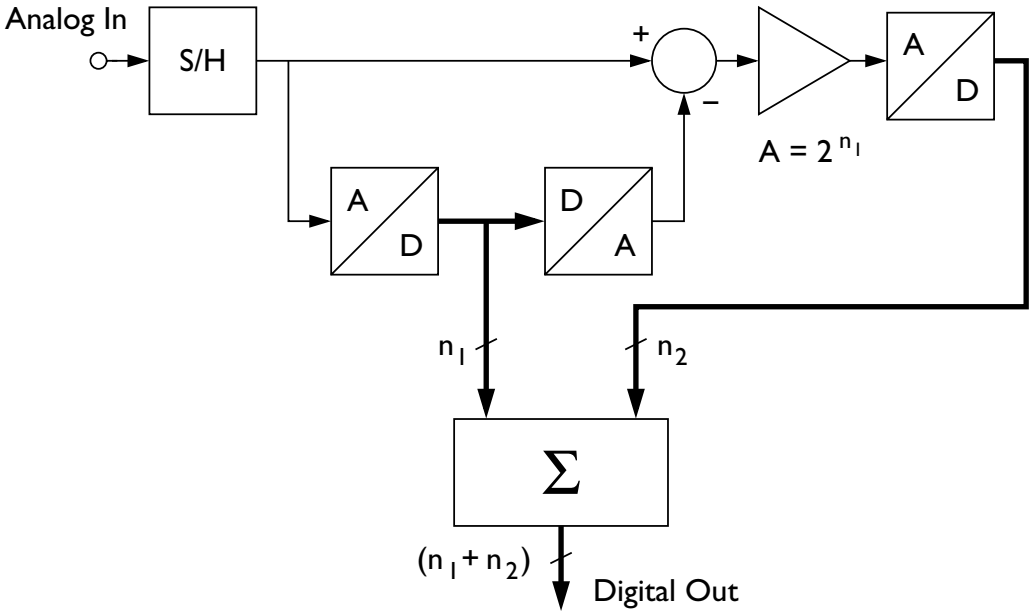

Figure 4.3: A two-step A/D converter

the drawback is evident. $(2^n - 1)$ comparators are necessary for an n bit flash converter. The exponential growth of the number of comparators and thus also silicon area and, more important, power dissipation, limits the resolution that can be achieved with a full flash A/D converter to about 10 bit. The resolution of most flash converters, however, does not exceed 8 bit.

4.2.2 Two-Step Architecture

4.2.2.1 Basic Two-Step Conversion

To increase the resolution but maintain high conversion rates, flash converters can be extended to a two-step or subranging architecture (also called series-parallel converter) [68, 69]. With this structure the conversion process is split into two steps as shown in figure 4.3.

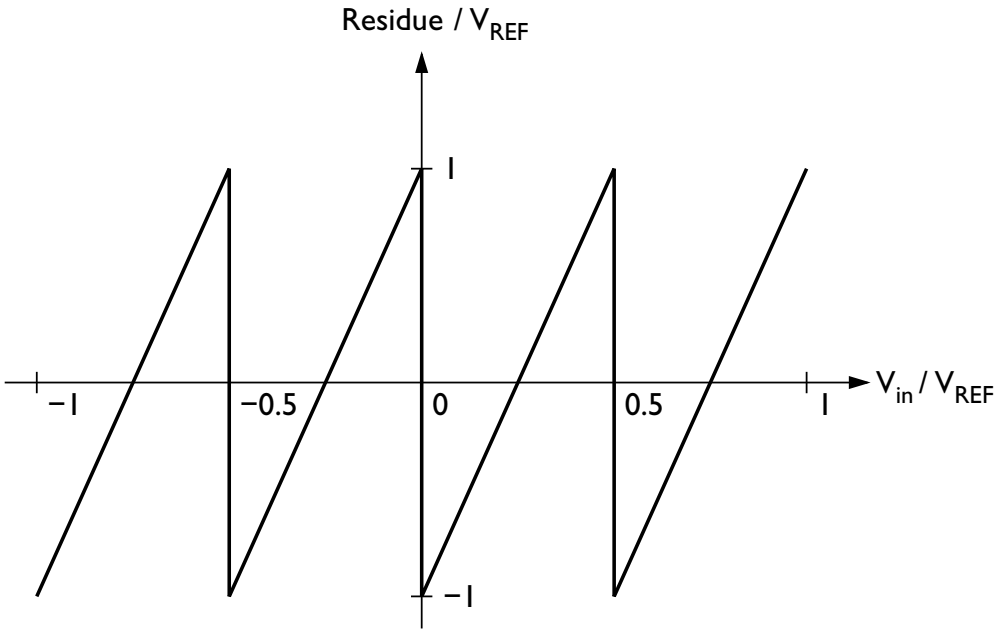

Figure 4.4: The amplified residue after a stage with a resolution of 2 bit

The first A/D subconverter performs a coarse conversion of the input signal. A D/A converter is used to convert the digital output of the A/D subconverter back into the analog domain. The output of the D/A converter is then subtracted from the analog input. The resulting signal, called the *residue*, is amplified and fed into a second A/D subconverter which takes over the fine conversion.[1] At the end of the conversion the digital outputs of both A/D subconverters are summed up.

The amplified residue of an ideal A/D converter with a first stage that has a resolution of 2 bit is shown in figure 4.4. The input range of

1 The amplification between the two stages is not strictly necessary, but is carried out nevertheless in most of the cases. With the help of this amplifying stage, the second A/D subconverter can work with the same signal levels as the first one, and therefore has the same accuracy requirements.

this converter is $-V_{\mathrm{REF}}$ to V_{REF}. The D/A converter output is shifted by 0.5 LSB to obtain a symmetrical transfer characteristic. This shift is necessary to obtain a residue which lies entirely in the conversion range of the subsequent stage. Beginning at the minimum input voltage ($V_{\mathrm{in}} = -V_{\mathrm{REF}}$), the residue rises with rising input voltage. At the moment the first decision level of the A/D subconverter is reached, the output of the converter changes from code 00 to code 01. This causes the D/A converter to subtract one LSB from the analog input which can be seen as a step in the residue. This behavior continues up to the maximum input voltage, creating the characteristic saw-tooth shape of the residue.

While the conversion speed of this architecture is only about half of the speed of a full flash type, the number of comparators decreases drastically. For an eight bit converter, e. g., a full flash implementation needs 255 comparators. A two-step architecture with four bits per stage, on the other hand, only needs 30 comparators, which is a reduction of silicon area and power consumption by a factor of 8.5.

4.2.2.2 Pipelined Operation

In contrast to the full flash A/D converter, the two-step architecture is equipped with a sample-and-hold (S/H) circuit in front of the converter (see figure 4.3). This additional circuit is necessary because the input signal has to be kept constant until the entire conversion (coarse and fine) is completed. By adding a second S/H circuit between the two converter stages, the conversion speed of the two-step A/D converter can be increased almost to the level of a flash converter. This circuit is depicted in figure 4.5.

In a first clock cycle, the input sample-and-hold circuit samples the analog input signal and holds the value until the first stage has finished its operation and the outputs of the subtraction circuit and the amplifier have settled. In the next clock cycle, the S/H circuit between the two stages holds the value of the amplified residue. Therefore, the second

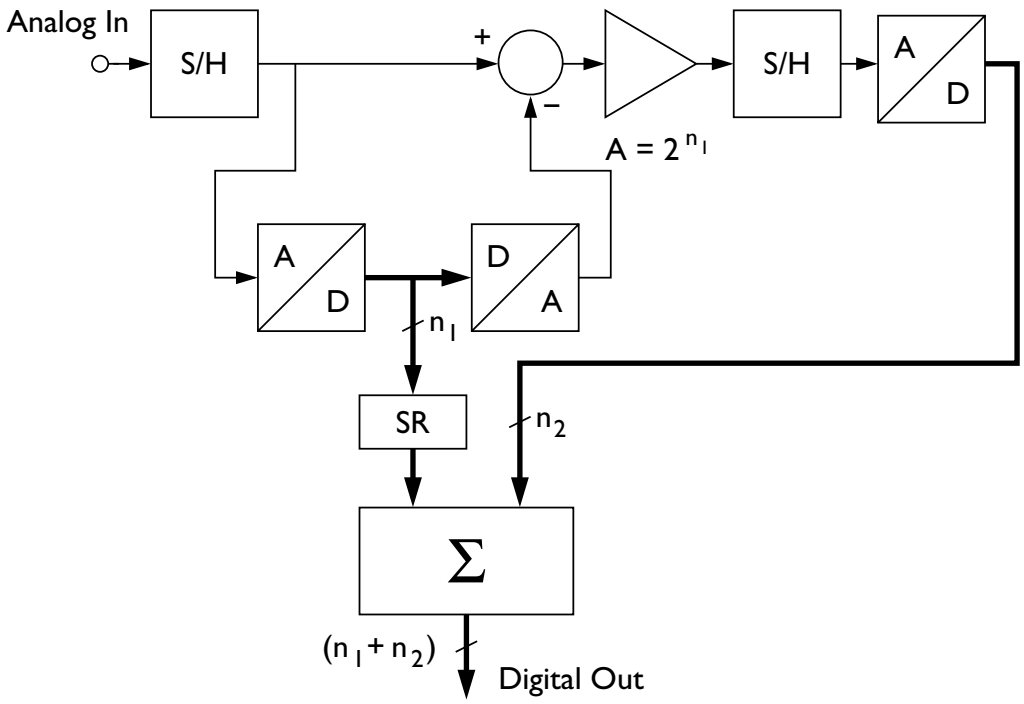

Figure 4.5: A two-step A/D converter with an additional sample-and-hold circuit and a shift register (SR) to line up the stage output in time

stage is able to operate on that residue independently of the first stage, which in turn can convert a new, more recent sample. The timing of this process is depicted in figure 4.6. The advantage of this so-called *pipelined* operation is the higher conversion speed compared to a two-step converter without an internal sample-and-hold circuit. The maximum sampling frequency of the pipelined two-step converter is determined by the settling time of the first stage only due to the independent operation of the two stages. To generate the digital output for one sample, the output of the first stage has to be delayed by one clock cycle by means of a shift register (see figure 4.5).

Although the sampling speed is increased by the pipelined operation, the delay between the sampling of the analog input and the output of the

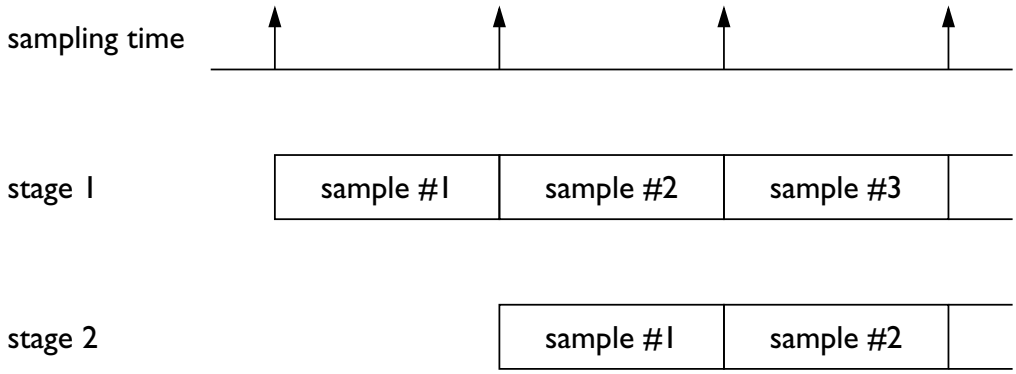

Figure 4.6: Basic timing of a pipelined two-stage A/D converter

corresponding digital value is still two clock cycles. The converter is said to have a *latency* of two clock cycles. For most applications, however, latency does not play any role, only conversion speed is important. In all signal processing and telecommunications applications, the main delay is caused by digital signal processing, so a latency of even more than two clock cycles is not critical.

4.2.3 General Pipelined Architecture

The architecture as described above is not limited to two stages. Because the interstage sample-and-hold circuit decouples the individual stages, there is no difference in conversion speed whether one single stage or an arbitrary number of stages follow the first one. This leads to the general pipelined A/D converter architecture, as depicted in figure 4.7 [6, 70–72]. The schematic of one pipeline stage is shown in figure 4.8.

The advantage of the pipelined A/D converter architecture over the two-step converter is the freedom in the choice of number of bits per stage. In principle, any number of bits per stage is possible, down to one single bit. It is even possible to implement a non-integer number of bits such as 1.5 bit per stage by omitting the top comparator of

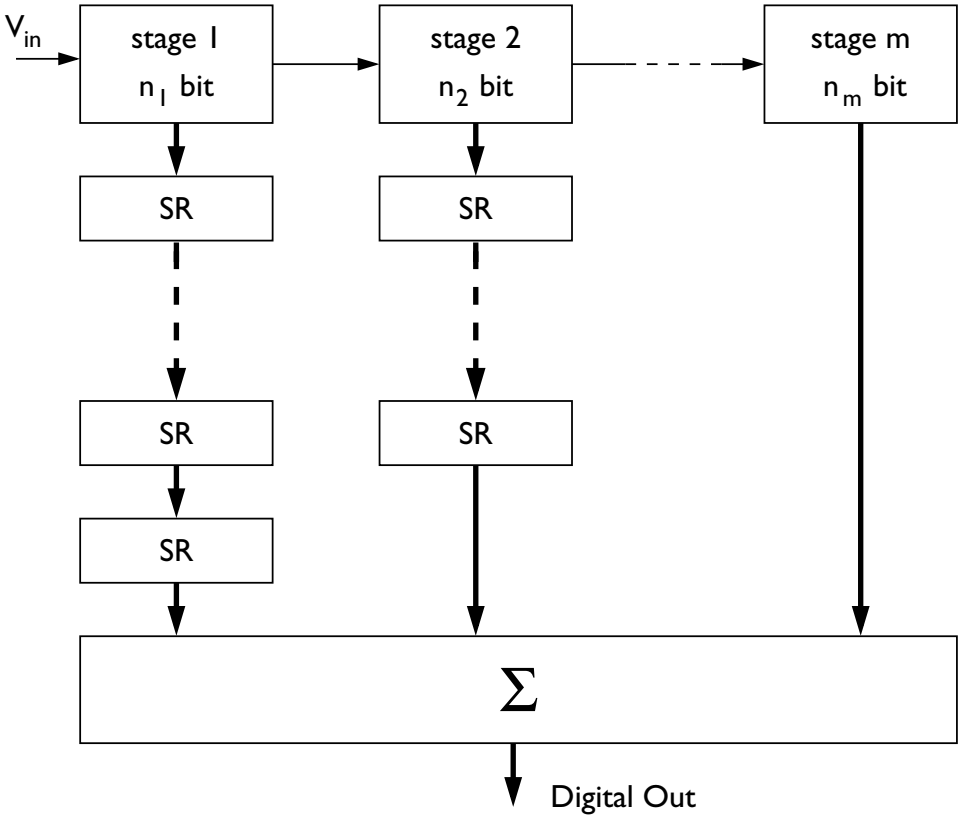

Figure 4.7: General block diagram of a pipelined A/D converter

the flash A/D subconverter used in the individual stages [72]. It is not necessary, although common, that the number of bits per stage is identical throughout the pipeline, but can be chosen individually for each stage [73, 74]. In particular for A/D converters with more than 12 bit resolution, the pipelined architecture is advantageous over a two-step solution because the latter requires a high number of bits per stage, which leads to the same problems (high power consumption and large silicon area) as discussed for the flash converter. The choice of the number of bits per stage for a particular application will be discussed in detail in section 4.4.

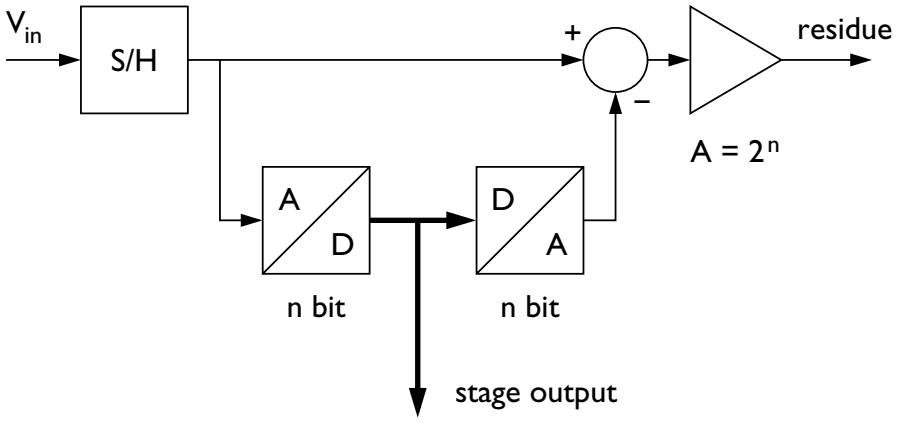

Figure 4.8: One stage of a pipelined A/D converter

The only real disadvantage of the pipelined architecture is the increased latency. For an A/D converter with m stages, the latency is m clock cycles. For architectures with a small number of bits per stage, the latency can thus be ten to fourteen clock cycles or even more. It depends on the particular application whether this is tolerable or not. Telecommunications systems, however, which have a strong need for high-speed A/D converters, are immune to latency, as discussed before.

4.3 Errors in Pipeline Stages

One stage of a pipelined A/D converter consists of several different building blocks as shown in figure 4.8. Not all, however, have the same influence on the accuracy of the entire A/D converter.

As discussed in section 4.2.2, the residue of a pipeline stage is generated by subtracting the D/A converter output from the analog input signal. The D/A converter therefore determines the accuracy of the residue, which contains the entire information about the portion of the

analog signal which has not been converted to the digital domain yet.[2] Therefore, the D/A converter has to provide the full accuracy of the complete A/D converter, although its number of output bits is the same as the stage resolution. In a 12 bit pipelined A/D converter with two bits per stage, e. g., the D/A converter of the first stage has only two output bits, but its accuracy has to be of a 12 bit level. Therefore, the performance of the entire pipelined A/D converter is limited by the linearity of the stage D/A converter.

4.3.1 A/D Subconverter Non-Linearity

Non-linearity of the A/D subconverter has a completely different effect on the performance of the entire A/D converter. The consequences of non-linearity can best be viewed by means of the residue. The residue in figure 4.9 is created by an A/D subconverter whose first decision level is shifted towards a lower value by 0.5 LSB compared to its ideal value, and its last decision level is shifted towards a higher value, also by 0.5 LSB.

Two things can be observed: first, the conversion range of the subsequent stage, which is also $-V_{\mathrm{REF}}$ to V_{REF}, is not sufficient any more to convert the residue without loss of information. Second, despite the fact that the A/D subconverter has a large non-linearity, the residue contains the entire information about the analog input signal. If the subsequent stage were able to convert this residue properly, no information would be lost. Moreover, the knowledge about the residue could be used to correct the error caused by the A/D subconverter of the first stage.

2 The conversion process of a pipelined A/D converter can be seen as follows. In the first stage, a few bits of information (i. e. the most significant bits) are extracted from the analog signal. The remainder of the information the analog input signal carries is fed to the second stage, where another portion of information is extracted. This goes on until to the last stage, where the digital output is completed. Only then the digital information is equal to the resolution of the A/D converter.

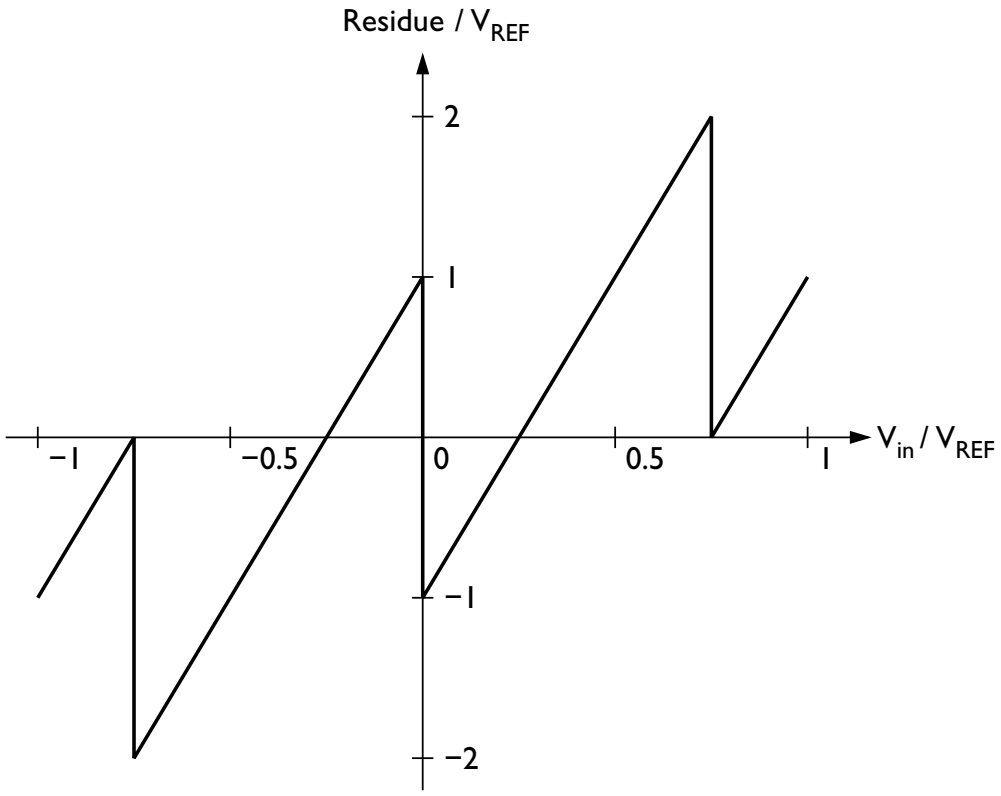

Figure 4.9: Residue of a pipeline stage having an A/D subconverter with large non-linearity

This corrective feedback is called *digital error correction* [6, 75]. As a first measure the gain of the interstage amplifier is reduced by a factor of two. This effectively doubles the conversion range of the subsequent stage. Therefore, residues like the one depicted in figure 4.9 can be converted without any loss of information. If the first stage is perfectly linear, however, only half of the conversion range of the second stage is used. Thus, one bit of the second stage is available to digitally correct the error of the first stage, the other bits add to the output of the pipelined A/D converter.

The correction of the output of the first stage is accomplished as follows. After the second stage has quantized the residue, the digital

output is evaluated. If the first stage is perfectly linear, only codes 01 or 10 result as the output of the second stage (assuming an architecture with two bits per stage). An overrange of the residue, i. e. the resulting code is 11, can only occur if the output of the first stage is too low. Therefore, one LSB has to be added to the output of the first stage. If the residue is underrange (indicated by a code 00 of the second stage), then one LSB has to be subtracted from the first stage output.

With the addition of one extra bit to the second stage, as described above, non-linearities of the A/D subconverter of the first stage up to ± 0.5 LSB can be corrected. More precisely, each decision level of the A/D converter can be shifted by as much as 0.5 LSB of the stage resolution in either direction without impairing the overall accuracy of the entire converter. In principle, it is also possible to use more than one bit of the second stage for digital error correction, but usually a correction range of ± 0.5 LSB of the stage resolution is enough.

The algorithm as described above has two disadvantages. First, to correct the digital output of one stage, addition and subtraction are necessary. Not only does this render the correction logic more complex than necessary, but it also presents difficulties when testing the functionality of the A/D converter [72]. Second, the detection of over- or underrange is not straightforward. This is of negligible importance if two bits per stage are used as in the example above, but adds to the complexity of the overall system if a larger number of bits per stage is used.

The solution is to eliminate the need of subtraction in the correction logic by adding offset to the A/D subconverter decision levels [72, 75]. If the decision levels of the A/D subconverter as well as the output values of the D/A converter are shifted towards more positive values by 0.5 LSB, the residue depicted in figure 4.10 results.

When comparing figure 4.10 to figure 4.4, the shift of the A/D subconverter decision levels is clearly visible. The first decision level, which ought to be at $-0.5\ V_{\mathrm{REF}}$ now resides at $-0.25\ V_{\mathrm{REF}}$, which is 0.5 LSB

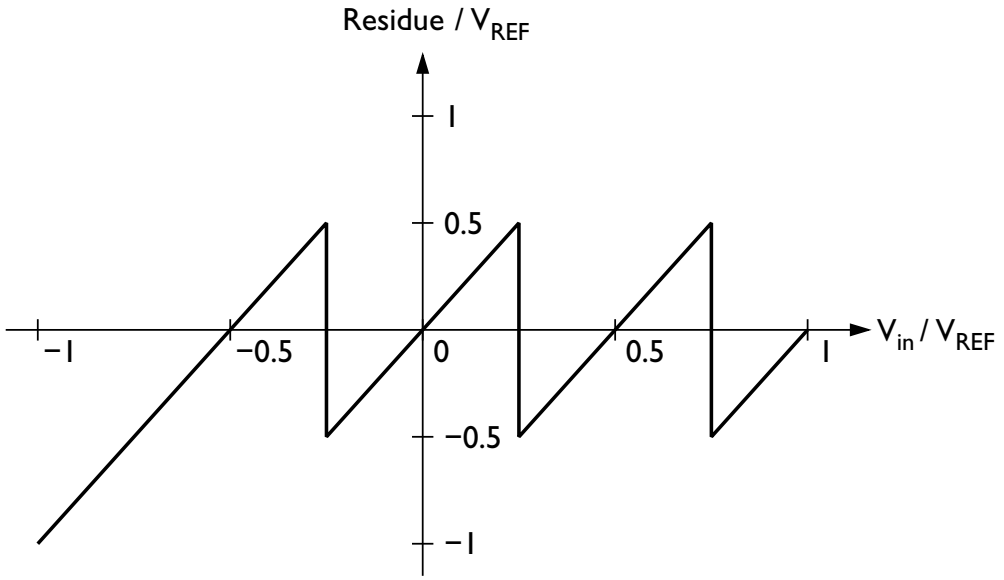

Figure 4.10: Amplified residue for addition-only digital error correction. The gain is already reduced in this figure.

higher than before. The offset of the D/A converter levels shifts the entire residue down. This results in an ideal range of the residue from $-V_{\mathrm{REF}}$ to 0. Any positive value of the residue is considered overrange by the error correction and will be adjusted as before. Therefore, error correction takes place even when the A/D subconverter of the first stage is ideal. This is caused by the deliberate offset of the subconverter decision levels. If non-linearity shifts these decision levels by no more than 0.5 LSB, only overrange of the residue can occur. This situation is depicted in figure 4.11, where the residue is caused by the same A/D subconverter errors as the residue shown in figure 4.9.

Due to the missing case of underrange in the residue, the digital error correction only has to perform addition of one LSB (if overrange occurs) or leave the digital output value unaltered. Overrange detection

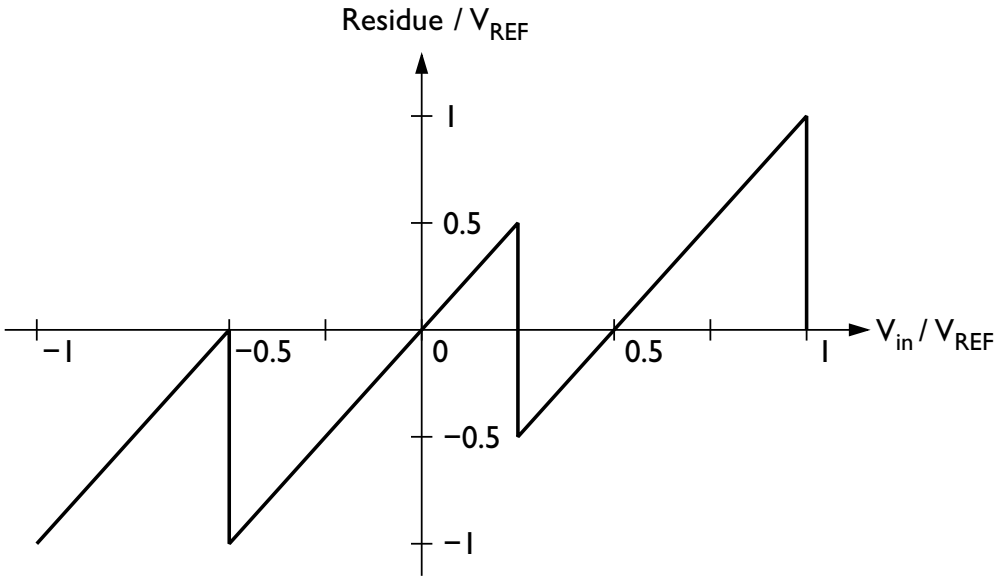

Figure 4.11: Residue with shifted decision levels for addition-only error correction and A/D converter non-linearity

is also easy because it only requires testing of the most significant bit of the output of the second stage.

The correction range of this *addition-only* error correction is identical to the original correction algorithm. Provided that one bit of the second stage is used for error correction, the correction range is ± 0.5 LSB.

Either algorithm of digital error correction has the advantage that the A/D subconverter in each stage is released from its accuracy requirements. Due to the corrective feedback, it is sufficient that the accuracy of the A/D subconverter only matches the stage resolution instead of the resolution of the entire pipelined converter. Furthermore, the speed of the entire converter is increased. The input sample-and-hold of each stage only has to settle to the stage resolution before the A/D subconverter can start its conversion. The full settling of the sample-and-hold output can take place simultaneously with the D/A converter settling [75].

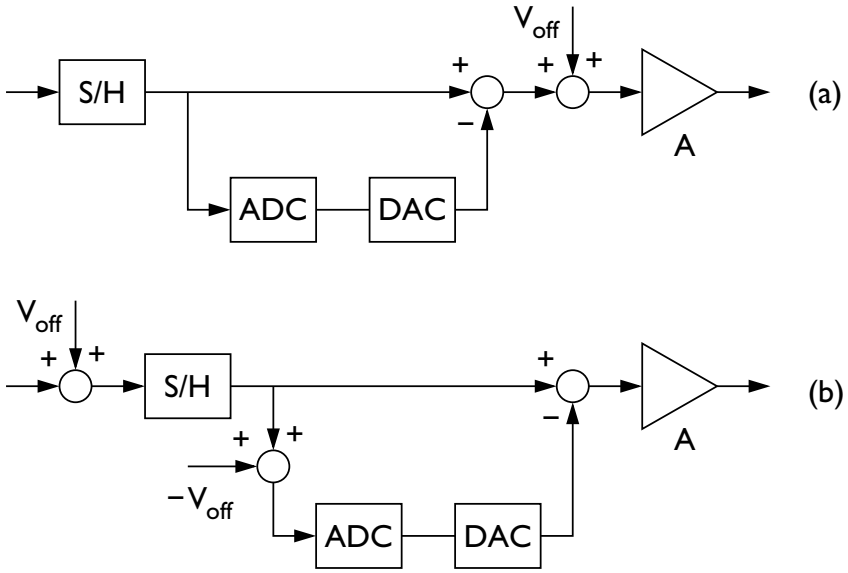

Figure 4.12: Equivalent offset voltage at the input of the gain stage (a), and the same offset moved towards the A/D converter input (b)

4.3.2 Offset Errors

Another source of errors which potentially degrades the performance of pipelined A/D converters is offset which occurs at the sample-and-hold stages and at the interstage amplifiers.

The offset error of the first sample-and-hold circuit at the input of the A/D converter only causes input-referred offset, but does not affect the linearity of the pipelined converter. The offset of the interstage amplifier and of the sample-and-hold circuit of a subsequent stage can be combined into one equivalent offset source at the input of the gain stage. This equivalent offset can be moved towards the input of the converter, as shown in figure 4.12. To obtain identical results, a second offset voltage of opposite polarity has to be inserted into the A/D subconverter branch [6]. Thus, the offset at the gain stage input is effectively split into two parts. The offset at the converter input presents no problems, as discussed above.

The offset in the A/D subconverter branch, on the other hand, can be corrected by the digital error correction presented in section 4.3.1. The offset is identical to a shift of *all* A/D subconverter decision levels, which can be corrected if the shift is smaller than ± 0.5 LSB. Therefore, any offset smaller than this amount can be corrected by the digital error correction if no other non-idealities occur. If the A/D subconverter suffers from non-linearity also, both errors can be corrected if their sum is smaller than ± 0.5 LSB.

4.3.3 Gain Errors

In addition to causing offset, the sample-and-hold circuits and the amplifiers also give rise to gain errors due to device mismatch. As in the case of offset, the gain error of the first sample-and-hold circuit only alters the conversion range of the pipelined A/D converter, but does not affect linearity.

The gain errors of the interstage amplifiers and the subsequent sample-and-hold circuits degrade the linearity of the pipelined converter. Moreover, if the interstage gain is smaller than the ideal value, missing codes can occur [70]. The influence of the gain error on the integral and differential non-linearities (INL and DNL, respectively) can be calculated as follows. It is assumed that the converter has a unipolar input range from zero to one and no error correction.[3] The gain of the interstage amplifier and the subsequent sample-and-hold circuit can be combined into one equivalent gain which can be expressed as

$$A = A_{\mathrm{id}}(1 + \Delta A) \tag{4.1}$$

where A denotes the equivalent gain, A_{id} is the ideal gain the amplifier and the sample-and-hold circuit should have, and ΔA is the gain error.

3 Because all errors are expressed in fractions of one LSB, the results obtained here also hold for A/D converters having a different conversion range than the one assumed here.

The stages following the amplifier that has non-ideal gain form an independent A/D converter which converts the remaining number of bits. Due to the non-ideal gain the code edges of this subsequent converter are shifted with respect to their ideal position. If no gain error is present, the ideal lower code edge $e_{i_{\mathrm{id}}}$ for code i of an n bit converter is given by

$$e_{i_{\mathrm{id}}} = \frac{i}{2^n} \qquad i = 0 \ldots (2^n - 1) \tag{4.2}$$

Any non-ideal gain shifts the code edges to the following position:

$$e_i = \frac{1}{1 + \Delta A} \frac{i}{2^n} \qquad i = 0 \ldots (2^n - 1) \tag{4.3}$$

The code centers c_i for each code i can be calculated from the two equations above. In the ideal case the code center $c_{i_{\mathrm{id}}}$ is given by

$$c_{i_{\mathrm{id}}} = \frac{2i + 1}{2^{n+1}} \qquad i = 0 \ldots (2^n - 1) \tag{4.4}$$

while the shifted code centers due to a gain error are located at

$$c_i = \frac{1}{1 + \Delta A} \frac{2i + 1}{2^{n+1}} \qquad i = 0 \ldots (2^n - 1) \tag{4.5}$$

The integral non-linearity is defined as the difference between the actual and the ideal code centers [5]. Using equations (4.4) and (4.5) and expressing the result in fractions of one LSB, the integral non-linearity for code i can be written as

$$\mathrm{INL} = \left(i + \frac{1}{2}\right) \left(\frac{-\Delta A}{1 + \Delta A}\right) \quad \mathrm{LSB} \tag{4.6}$$

The peak value of the INL occurs when the magnitude of the residue is maximum. Because the INL is not defined for the outermost codes,

the peak INL occurs at 1 LSB below the maximum code. Using equation (4.6) the peak INL therefore equals

$$\mathrm{INL}_{\mathrm{peak}} = \left(\frac{3}{2} - 2^{n_s}\right)\left(\frac{\Delta A}{1+\Delta A}\right) \quad \mathrm{LSB} \tag{4.7}$$

where n_s denotes the number of bits of the A/D converter formed by the stages following the non-ideal amplifier.[4] For small values of the gain error ΔA, equation (4.7) can be simplified to

$$\mathrm{INL}_{\mathrm{peak}} \approx -\delta \ \mathrm{LSB} \quad \text{for} \quad \Delta A = \delta\, 2^{-n_s} \tag{4.8}$$

The differential non-linearity is defined as the deviation of the step width of each code from one LSB [5]. Inspecting equations (4.2) and (4.3) shows that the code edge of the highest code, which is code $(2^n - 1)$, is shifted by the largest amount while the code edge of code 0 is not shifted at all. Therefore the peak DNL error is expected to occur at the code bounded by these two code edges. Using equation (4.3) to determine the actual code width and subtracting one LSB leads to the following result:

$$\mathrm{DNL}_{\mathrm{peak}} = 2^{n_s}\left(1 - \frac{1-2^{-n_s}}{1+\Delta A}\right) - 1 \tag{4.9}$$

This expression, too, can be simplified for small values of ΔA:

$$\mathrm{DNL}_{\mathrm{peak}} \approx \ \delta \ \mathrm{LSB} \quad \text{for} \quad \Delta A = \delta\, 2^{-n_s} \tag{4.10}$$

From equation (4.8) the required gain accuracy for a total integral non-linearity of ± 0.5 LSB can be written as [76]:

$$|\Delta A| \leq \frac{1}{2^{(n_s+1)}} \tag{4.11}$$

4 If digital error correction is applied, n_s denotes the number of bits which are used for conversion. The bits used for correction have to be subtracted.

Missing codes occur if the gain error ΔA is negative and its absolute value is bigger than 2^{-n_s}. For positive values of ΔA no missing codes occur.

If an A/D converter with bipolar input range from $-V_{REF}$ to V_{REF} is used, the maximum of the residue is only half as big as in the unipolar case (see figure 4.4). Therefore, the gain inaccuracy can be twice as high as in the unipolar case to produce the same error. Equation (4.11) can be relaxed to

$$|\Delta A| \leq \frac{1}{2^{n_s}} \tag{4.12}$$

The INL and DNL given in equations (4.7) and (4.9) also have to be divided by a factor of two. Moreover, the prediction in these two formulas is slightly too large due to the fact that the maximum occurs two LSB below the maximum code for the bipolar case.

The equations above can be interpreted as follows. As long as the error created by the inaccurate gain is smaller than ± 0.5 LSB of the combined resolution of the following stages, the integral non-linearity of the entire converter will also be below ± 0.5 LSB, which means that the gain error does not affect the accuracy of the pipelined converter.

Equation (4.11) shows that the gain after the first stage of the A/D converter has to be most accurate, because the number of remaining bits n_s is maximum after the first stage. The amplifiers in the subsequent stages have relaxed accuracy requirements. Gain accuracy is of special importance for pipelined converters with high resolution which utilize many stages with a low number of bits. The interstage gain accuracy after the first stage of a 14 bit A/D converter, e. g., has to be equivalent to 12 bit resolution if only two bits per stage are converted.

Gain errors cannot be corrected by digital error correction [6, 72]. While the original error correction algorithm that executes addition and subtraction does not alter the residue and thus does not influence the errors caused by inaccurate gain, addition-only error correction causes an increase of the gain-induced linearity errors [72]. The reason is the shifting of the A/D subconverter and D/A converter levels which in-

creases the absolute value of the residue for small input values (cf. figure 4.10). Due to the larger residue the non-linearity caused by a gain error increases. In practice, this disadvantage is small. Every linearity error in the A/D subconverter increases the residue and thus worsens gain error effects. Therefore, the increase in non-linearity due to the increased residue only applies to ideal A/D converters. For real implementations the performance of the two error correcting algorithms is almost identical.

4.4 Choosing the Stage Resolution

The knowledge about errors in the pipeline stages that affect the linearity of the A/D converter is the basis upon which the resolution of the stages is chosen. For subranging architectures that do not feature an interstage sample-and-hold circuit the conversion speed decreases with increasing number of stages because all stages have to settle before the next conversion cycle can start. In contrast, the pipelined architecture decouples the individual stages such that the conversion speed is almost independent of the number of stages, as was pointed out in section 4.2.3.

Due to this independence between the number of stages and the conversion speed, the resolution of each stage can be chosen to satisfy other requirements than speed. The choice of stage resolution influences the linearity, speed, area and power consumption of the pipelined A/D converter [76].

The linearity of the A/D converter depends on the stage resolution in two ways. First, all errors that are caused by subsequent stages are attenuated by the interstage gain when referred to the input of the A/D converter. An error caused by the second stage is attenuated by the gain of the first amplifier, while errors stemming from the third stage are attenuated by the gain of two amplifiers already. Thus, if the gain between the individual stages is greater than unity, the errors of the first

stage will dominate. This in turn means that for linearity reasons the stage resolution should be chosen larger than one bit if digital error correction is applied. This leads to a minimum stage resolution of two bit, or 1.5 bit for certain error correction algorithms [72, 76]. More bits per stage attenuate errors stemming from subsequent stages even more, but a gain of two is usually sufficient.

The second mechanism how the stage resolution affects the overall linearity of the A/D converter is by the gain error of the interstage amplifier. As discussed in section 4.3.3, the gain of the amplifiers between the stages must be accurate enough to preserve the linearity of the subsequent stages. Most important in this context is the interstage gain after the first stage because the remaining resolution and thus the necessary gain accuracy is maximum after the first stage. The higher the resolution of the first stage, the more relaxed the gain requirements of the first interstage amplifier are. Therefore, from a linearity point of view, a large stage resolution, at least for the first stage, is preferable.

The conversion speed, unfortunately, has other requirements. For a given unity-gain bandwidth of the amplifier, which is limited by the technology that has been chosen for the implementation, the settling time of the gain stage is minimized if the gain and thus the stage resolution is as small as possible. This is the case even though the settling accuracy has to be higher for a low stage resolution due to the larger number of remaining bits after that particular gain stage [76]. Therefore, for fast A/D converters the stage resolution has to be as small as possible.

Power consumption and silicon area also depend on the stage resolution. While silicon area is usually not of much concern, power consumption is. Estimations of the dependency of power on the stage resolution are less straightforward than linearity and speed are. For medium resolution A/D converters, a minimum stage resolution also seems to minimize the power consumption [76]. For high resolution converters, on the other hand, a higher number of bits per stage is preferable from a power consumption point of view [12, 77].

Summarizing the process of choosing the stage resolution for a pipelined A/D converter, the best overall compromise for medium resolution converters up to 10 bit is a low number of bits per stage [72, 76]. This has led to a large number of designs using two bits per stage or even 1.5 bit per stage [64, 78–80].

For higher resolutions, however, the disadvantages of a low stage resolution on the converter linearity become more important. In that case, increasing the stage resolution helps to preserve the accuracy of the pipelined A/D converter. Therefore, many designs having a resolution of 12 bit or more use a stage resolution of 3 bit to 5 bit [12, 15, 73, 77, 81]. Another possibility to preserve the speed advantages of low stage resolution without compromising the converter linearity is to use a higher resolution only in the first stage where the gain error is most critical [74].

4.5 Summary

This chapter gives an overview of possibilities to increase the conversion speed of an A/D converter, starting from the design presented in chapter 3. The pipelined architecture is found to be well suited for A/D converters with an arbitrary number of bits and high conversion rates.

An in-depth understanding of the pipelined architecture is the basis for the design process and the compromises that have to be made during the design. Especially the errors stemming from the individual building blocks of a pipeline stage have to be analyzed and determined how they affect the performance of the entire A/D converter.

Digital error correction, which is a corrective feedback using redundant digital data, is able to correct linearity errors of the stage A/D converters as well as offset errors. Moreover, it accelerates the operation of the pipeline by postponing the full settling of the sample-and-hold amplifiers. Other errors, such as D/A converter non-linearity and gain errors in the amplifiers between the individual stages, cannot be cor-

rected. These errors, as well as speed requirements influence the choice of the resolution of the individual pipeline stages.

This chapter discusses all the architectural issues of the pipelined A/D converter that is presented in the next chapter. So the following chapter can concentrate on the particular implementation of this converter, emphasizing the use of the MOSFET-only ladder.

5

A Pipelined 13 bit A/D Converter

Du wähnst, es füge sich sogleich;
Hier stehen wir vor steilern Stufen.

MEPHISTOPHELES, Der Tragödie zweiter Teil

The final step in the investigation of A/D converters described throughout this thesis is to show that the MOSFET-only ladder not only is capable of good performance in medium speed applications, but also in A/D converters with higher conversion rates.

It has been pointed out in the discussion of chapter 4 that the pipelined architecture is the best solution to increase the conversion speed for this type of application. Moreover, the relatively small resolution of one stage in a pipelined A/D converter is advantageous for MOSFET-only ladders. Due to the small number of bits the D/A converter is very compact which increases the matching between the individual transistors of the ladder.

The speed of the ladder almost exclusively depends on the channel length of the MOS transistors, which is demonstrated in equation (2.37) of section 2.4. To reduce the settling time and thus increase the conversion speed it is therefore mandatory to use shorter channels. If the matching accuracy is to be maintained, however, the only solution is the choice of a technology with finer lithography. Therefore a 0.25 µm CMOS process has been selected for the implementation. The low supply voltage of 2.5 V that accompanies a 0.25 µm process is a disadvantage that has to be accepted. Its influence on the design will be discussed in more detail along with the description of the individual building blocks.

The measurements of the successive-approximation A/D converter presented in chapter 3 show that the accuracy of the MOSFET-only ladder can be 12 bit or more. Therefore, the resolution of this pipelined A/D converter has been chosen as 13 bit. By increasing the conversion rate to several Msample/s, this A/D converter fulfills the requirements for new wire-bound telecommunication systems such as ADSL (Asymmetric Digital Subscriber Line) or VDSL (Very High Speed Digital Subscriber Line). As discussed in chapter 1, these applications are currently the driving force behind A/D converter development.

5.1 Architecture

One of the most important architectural decisions is the stage resolution. Following the guidelines discussed in section 4.4, a stage resolution of four bit has been chosen for this A/D converter. The main reason for this decision is the gain accuracy of the interstage amplifier after the first stage. As will be described below, the gain stage consists of an operational amplifier with resistive feedback. After the four bits of the first stage, the accuracy requirement of the interstage gain is 9 bit. This matching accuracy normally can be achieved with polysilicon resistors without problems.

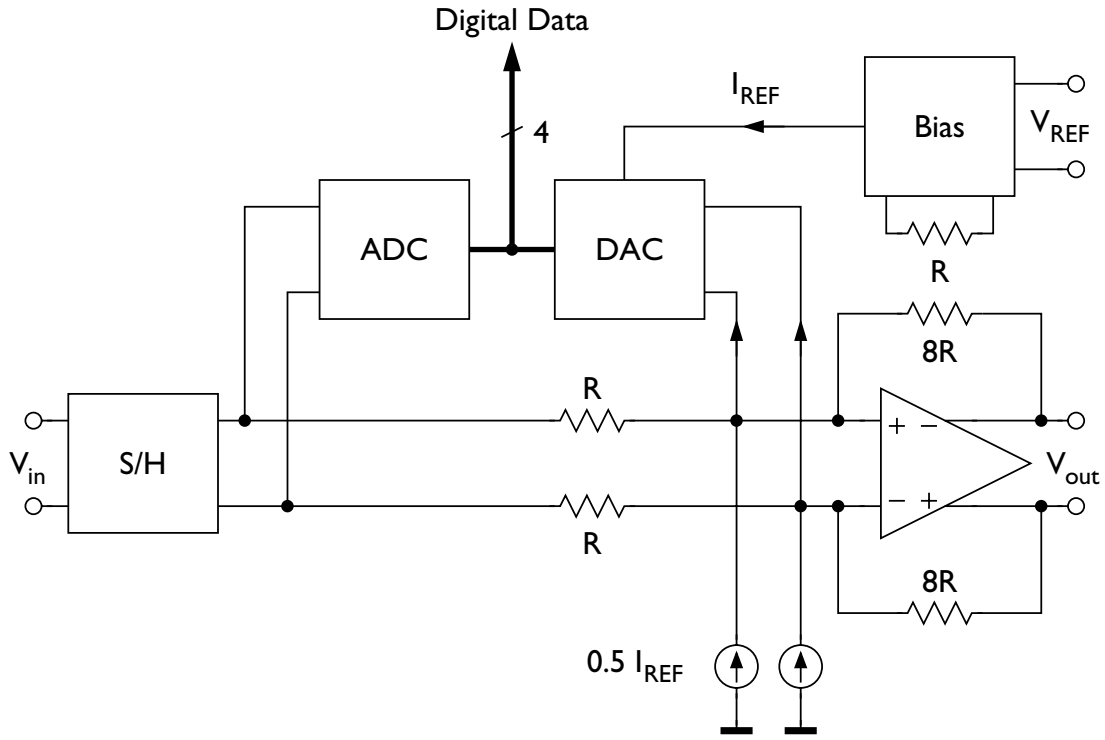

Figure 5.1: Extended block diagram of one pipeline stage

The block diagram of one pipeline stage is depicted in figure 5.1. Compared to the schematic of a traditional pipelined converter (without a MOSFET-only ladder), some differences exist. The most noticeable one is the separation of the gain stage and the sample-and-hold circuit. In most of today's implementations these two functions are combined into one sample-and-hold amplifier between the individual stages, only the input S/H is a block of its own [6]. The reason for two separate building blocks in this A/D converter is twofold. If a sample-and-hold circuit with gain is to be implemented, accurately matching capacitors are needed. As mentioned above, the matching accuracy has to be at least 9 bit after the first stage, which cannot be guaranteed to be achievable with sandwich capacitors consisting of five to six layers of metal.

The second reason is the output signal of the D/A converter which is a current instead of a voltage. The pros and cons of using currents versus voltages as signals have been discussed in section 3.2. While current was the signal of choice in the successive-approximation A/D converter, voltage is preferable here. The reason is the sample-and-hold circuit between the stages, which can be implemented more easily and accurately in the voltage domain than in the current domain. The consequence of this decision is that the output current of the D/A converter has to be converted back to a voltage before the sample-and-hold stage.

The implementation of the interstage gain using an operational amplifier with resistive feedback, as depicted in figure 5.1, performs the current-to-voltage conversion in an elegant manner. Before this conversion takes place, the necessary subtraction of the D/A converter output from the analog input signal (cf. figure 4.8), which can be done more accurately in the current domain, is accomplished at the virtual ground nodes of the amplifier. The current difference is then converted back to the voltage domain by the feedback resistors.

The only drawback of this architecture is the fact that the output voltage of the D/A converter generated in the way described above depends on the absolute value of the feedback resistors which can vary by $\pm 20\%$. This amount of variation is clearly not acceptable. The solution to this problem is to generate the reference current I_{REF} for the D/A converter on-chip by using external reference voltages and an internal resistor that is matched to the feedback resistors. In this way the output voltage of the D/A converter always matches the reference voltage, regardless of the absolute values of the resistors. Only the reference current changes by the same amount as the absolute value of the resistors does. This, however, is not critical because the speed of the ladder is independent of the value of the reference current (cf. section 2.4).

Another important aspect of the overall performance of a pipelined A/D converter is the timing of the individual pipeline stages. The basic timing of a pipelined A/D converter with four stages is depicted in

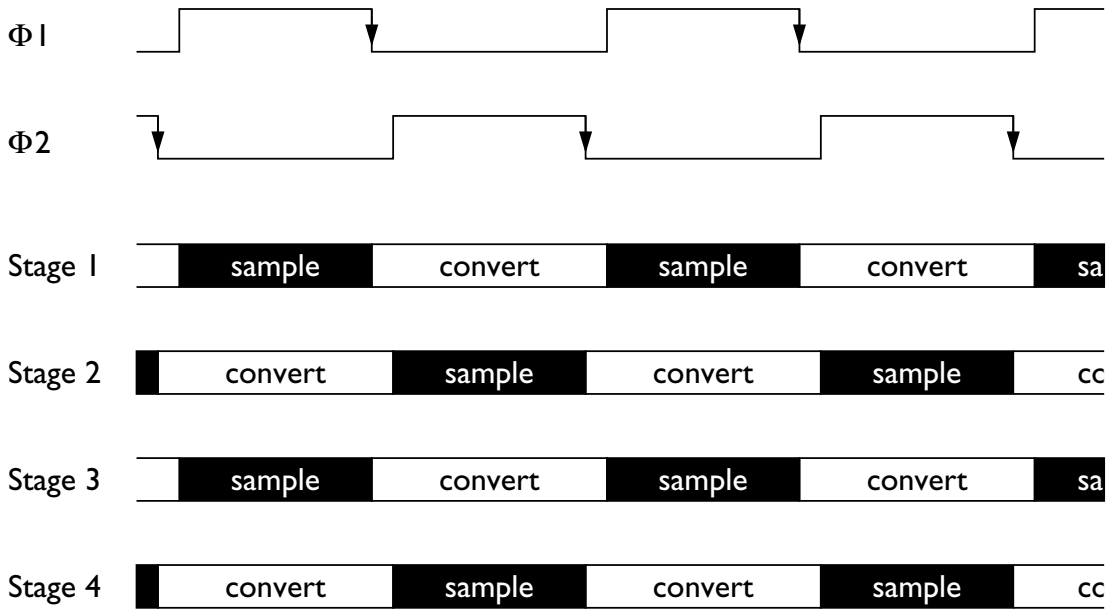

Figure 5.2: Basic timing of a pipelined A/D converter showing the interaction between the individual stages

figure 5.2 [8, 71]. The alternate operation of the individual pipeline stages is clearly visible. In the clock phase where the first stage is in normal operation, i. e. when it converts the input and amplifies the residue, the second stage is in sample mode. At the end of this clock phase, the second stage samples the output residue of the first stage (indicated by an arrow in figure 5.2) and holds it during the next clock phase when the second stage performs its conversion.

This timing scheme is suitable for converter architectures where the time needed for the sampling (i. e. the acquisition time of the sample-and-hold circuit) is approximately the same as the conversion and settling time of the pipeline stage. This is not the case, however, in the implementation described here using the MOSFET-only ladder. While the acquisition time of the sample-and-hold stage can be made very fast due to the small feature size of the technology, the settling time of the

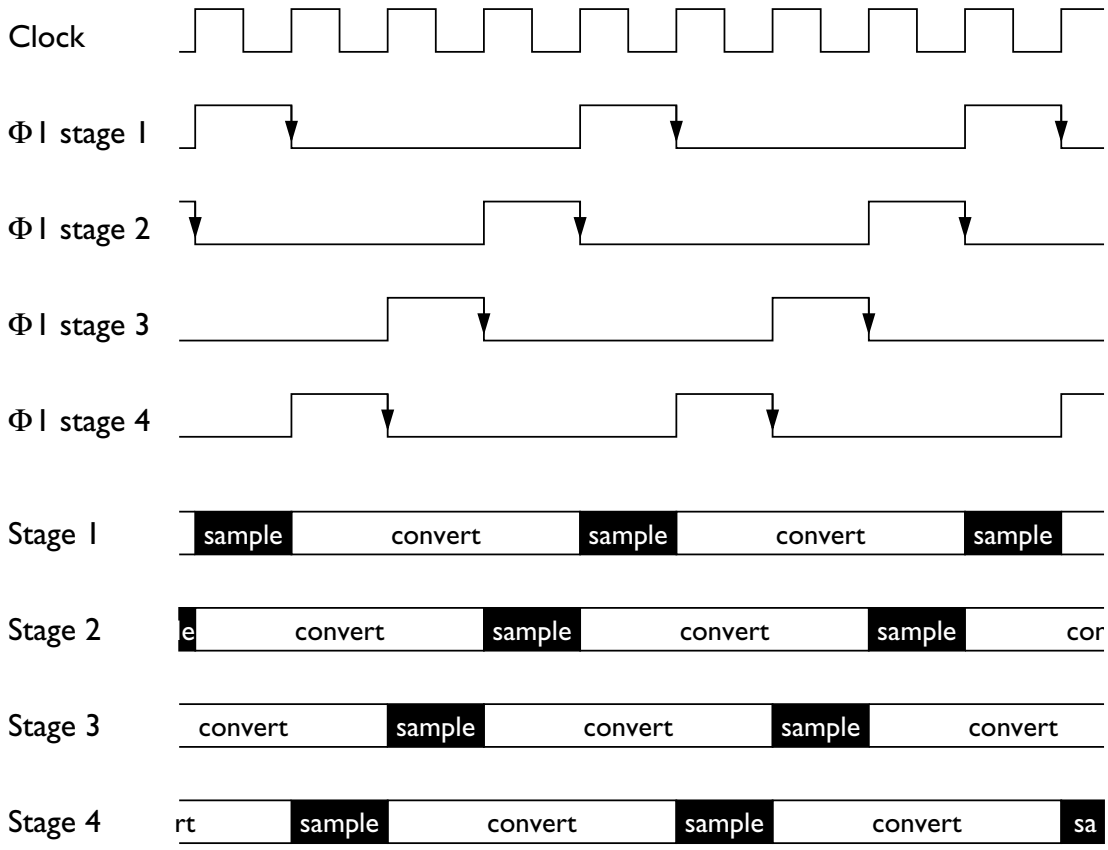

Figure 5.3: Timing of the pipelined A/D converter. Only clock phase Φ1 is shown for each stage. Non-overlapping of the clocks is also ignored.

D/A converter is limited by the minimum acceptable channel length of the ladder transistors which is dictated by matching considerations.

To obtain maximum conversion speed, the lengths of the sample phase and the conversion phase, respectively, have been made different, as shown in figure 5.3. The individual stages have shifted clock signals which are generated from a single external clock. The clock frequency is four times the sampling frequency. Although the clocking scheme is more complex than the basic case depicted in figure 5.2, the operation is exactly the same: At the end of the conversion phase of the first stage

the sample-and-hold circuit of the second stage samples the amplified residue. The only difference is that the sample-and-hold circuit of the second stage is in sample mode only for the last third of the conversion phase of the first stage.

Apart from being able to balance the settling time requirements of the sample-and-hold circuit and the D/A converter and gain stage, the generation of the clocks for the individual stages from a faster clock signal has the advantage that special clock phases for auto-zeroing and other auxiliary functions can be generated easily.

One last decision that has to be made before the individual building blocks of the pipeline stages can be designed is the choice of the digital error correction algorithm. This decision affects the A/D subconverters as well as the D/A converter. The addition-only digital error correction has several advantages from a digital point of view, as discussed in chapter 4. The necessary shifting of the A/D subconverter decision levels on the one hand and of the output values of the D/A converter on the other hand can be implemented without additional cost and without compromising the accuracy of these two building blocks. Therefore the benefits of easier implementation are the decisive factor to choose addition-only digital error correction.

The individual building blocks of the pipeline stage, as depicted in the block diagram of figure 5.1, are described in more detail in the following sections. The necessary specifications of the building blocks to obtain the desired overall accuracy as well as the circuit implementations that result from these specifications are discussed.

5.2 D/A Converter

The discussion of the block diagram of one pipeline stage in the last section shows the strong influence the usage of the MOSFET-only ladder in the D/A converter has on the architecture of the entire pipeline stage.

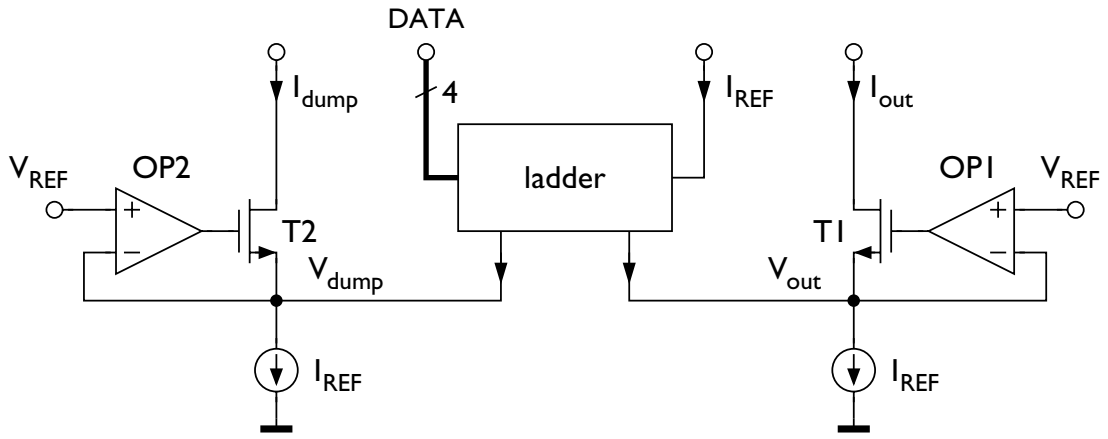

Figure 5.4: Block diagram of the D/A converter using the MOSFET-only ladder

The detailed implementation of the D/A converter also subordinates itself to the requirements of the ladder to obtain maximum linearity. The block diagram is depicted in figure 5.4.

5.2.1 Incorporating the Ladder

The extraction of the ladder output current is performed in the same way as in the successive-approximation A/D converter of chapter 3, namely by using a regulated cascode structure. In principle, the input terminals of the operational amplifier that implements the interstage gain also could have been used as a summing point for the ladder currents (cf. figure 5.1). The addition of a separate regulated cascode summing stage as shown in figure 5.4, however, has two major advantages. If the feedback gain stage would be used to perform the current extraction, the operational amplifier would be subject to high gain requirements to generate the low input impedance that is necessary for the ladder, as discussed in section 3.2. The minimum required gain of the operational amplifier for the interstage gain block would be 104 dB as opposed to the implemented solution with a separate current extraction stage, where a minimum gain of 71 dB is sufficient (see section 5.2.3).

The second advantage is the complete decoupling of the ladder output from the interstage gain block. Due to this separation the settling of the voltage at the interstage amplifier input, which is caused by a step at the sample-and-hold output or a change in the ladder current, does not affect the voltage at the ladder output terminals. Therefore, the settling of the interstage amplifier does not affect the settling of the ladder output current. This separation leads to a significant decrease of the overall settling time of the pipeline stage.

The dimensions of the ladder are chosen in accordance with the guidelines given in section 2.5. As mentioned at the beginning of this chapter, a 0.25 µm technology has been selected for the implementation of the pipelined A/D converter to be able to decrease the size of the ladder transistors. The channel length L is 5 µm, which enables operation at a 5 MHz clock while maintaining the necessary matching accuracy. The corresponding channel width is 22 µm. The reference current has been selected as 500 µA which yields an LSB current of 61 nA at a 13 bit level. Although this is a small value, it is still several orders of magnitude larger than the sum of the junction diode leakage currents. A larger reference current, on the other hand, would only increase the power consumption of the converter without increasing the ladder speed.

5.2.2 Current Extraction

The ladder dimensions as stated above result in a minimum output resistance of 1.5 kΩ. According to equation (2.27), the maximum allowed offset voltage between the nodes V_{out} and V_{dump} in figure 5.4 is 45 µV. With this value the output current error stays below 0.5 LSB.

This small offset voltage level cannot be accomplished without special circuitry, even when large transistor geometry is used. Therefore, active offset suppression such as auto-zeroing has to be performed [82]. During an auto-zero phase the input offset is stored on a capacitor. In the operation phase this capacitor is connected in series with the input and thus cancels the offset. Charge injection of the necessary switches,

however, causes a pedestal on the storage capacitor and thus results in residual offset. This residual offset can be reduced by either using a larger storage capacitor or by decreasing the size of the switch transistors. The speed requirements, however, put a limit on capacitor size and on the switch resistance and therefore on the ability to decrease the pedestal. To further reduce the residual offset, special countermeasures such as dummy switches, which cause identical charge injection as the main switch, but with opposite polarity, can be used [83].

The reduction of residual offset using dummy switches is limited by the matching accuracy between the main switch and the dummy switch [84]. An offset reduction to the value needed for the MOSFET-only ladder in this implementation, however, is unlikely to be achieved with this method, especially when the hold capacitor must be comparatively small to meet the speed requirements.

To further improve the offset performance, additional gain can be used to attenuate the error caused by charge injection [85]. The regulated cascode including this type of auto-zero scheme is depicted in figure 5.5. The operational amplifier has an additional auxiliary input. The gain from the main input to the output is A_{main}, while the auxiliary input has a reduced gain A_{aux} to the output. Due to the different gain values of the main and the auxiliary inputs the offset of the amplifier is stored on the hold capacitor amplified by the ratio $A_{\text{main}}/A_{\text{aux}}$. When switching back to normal operation, the charge injection is attenuated by the same factor. The detailed operation is as follows. During the first clock phase $\Phi 1$ the main inputs of the operational amplifier are short-circuited, and the regulated cascode loop is closed via the auxiliary input. Ignoring the output voltage of the operational amplifier (i. e. assuming very high gain), the voltage V_{c} across the hold capacitor C_{H} is equal to

$$V_{\text{c}} = \frac{A_{\text{main}} V_{\text{off}}}{1 + A_{\text{aux}}} \tag{5.1}$$

where the reference voltage V_{REF} is assumed to be zero for convenience. When the amplifier is switched back to normal operation in clock phase

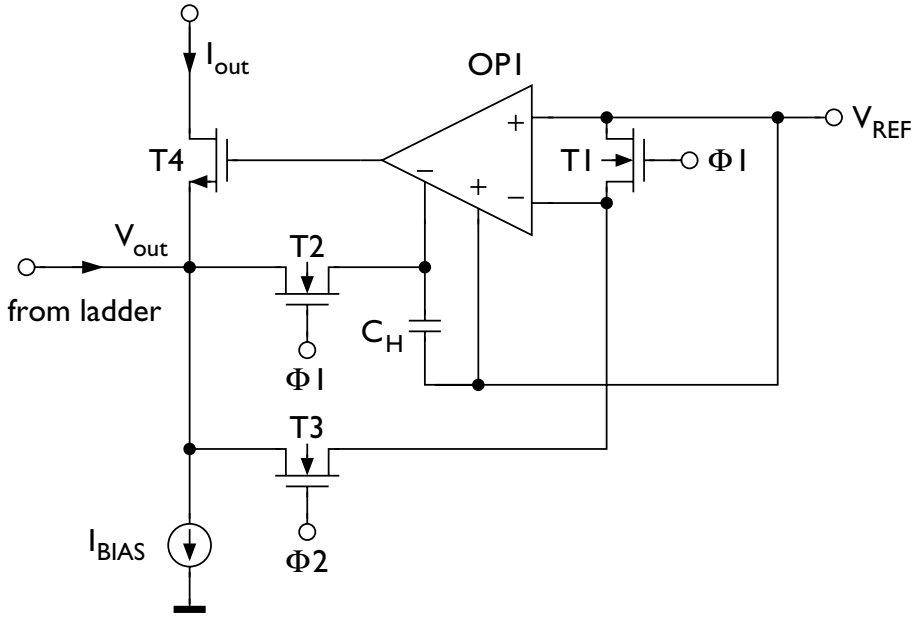

Figure 5.5: Regulated cascode stage with auto-zero using additional gain to reduce the errors induced by charge injection

Φ2, the regulated cascode feedback loop is closed through the main input of the operational amplifier. The amplified offset that is applied to the auxiliary input is subtracted from the output, attenuated to its original value due to the reduced gain A_{aux}. One important point of this auto-zero scheme is the fact that the feedback loop is closed in both clock phases. This is necessary to prevent the output of the operational amplifier from leaving the range of linear operation and going into saturation. If this happens, a long time is required for recovery before the feedback loop is operational again. The voltage V_{out} that appears at the ladder output terminal is given by

$$V_{\text{out}} = V_{\text{REF}} + \frac{V_{\text{off}}}{1 + A_{\text{aux}}} + \Delta V_{\text{inj}} \frac{A_{\text{aux}}}{A_{\text{main}}} \tag{5.2}$$

where ΔV_{inj} denotes the voltage error caused by charge injection due

to switch *T2*. Equation (5.2) shows that the pedestal caused by charge injection when switch *T2* is switched off is attenuated by the ratio of the two gains. It also shows that the original offset voltage V_{off} is attenuated by the gain of the auxiliary input, which presents a lower bound on the gain of the operational amplifier. A large ratio $A_{\text{main}}/A_{\text{aux}}$ is desirable to maximize the attenuation of errors caused by charge injection, but other requirements limit this gain ratio. The amplified offset is stored on the hold capacitor. Therefore, the gain ratio has to be chosen such that the maximum offset that is likely to occur still can be stored on the capacitor with the given voltage headroom. Speed requirements put another limit on the gain ratio. The amplifier from the auxiliary inputs to the output not only has lower gain than the main amplifier, but its unity-gain bandwidth is also reduced by the factor $A_{\text{main}}/A_{\text{aux}}$ compared to the main amplifier. If the gain ratio is chosen too large, excessive bandwidth is necessary for the main amplifier. The best compromise in this design is a gain ratio of eight. The suppression of the charge injection is large enough to meet the required offset of 45 µV, while the resulting unity-gain bandwidth of the main amplifier is still manageable.

Due to the low reference voltage V_{REF} of 0.4 V NMOS transistors can be used as switches. The most important switch is *T2* which causes the charge injection error when it is switched off at the end of clock phase Φ1. To reduce the original, non-attenuated charge injection, a dummy switch is provided. Although the size of the switch should ideally be as small as possible, the transistor dimensions are dictated by speed and stability requirements.

The on-resistance of the switch *T2* forms a pole together with the hold capacitance. Along with the dominant pole of the operational amplifier, a two-pole system is created whose stability must be examined. In the closed-loop system, the two poles are located at

$$p_{1,2} = -\frac{1}{2}\,\omega_{\text{s}}\left(1 \pm \sqrt{1 - \frac{4\,\omega_{\text{GBW}}}{\omega_{\text{s}}}}\right) \tag{5.3}$$

with

$$\omega_s = \frac{1}{r_{on} C_H} \tag{5.4}$$

where ω_s denotes the pole caused by the switch, ω_{GBW} is the gain-bandwidth product of the amplifier when using the auxiliary input, and r_{on} denotes the on-resistance of switch *T2*.

The optimum speed is achieved if both poles are real and identical.[1] This situation occurs if the following equality holds:

$$\omega_s = 4\,\omega_{GBW} \tag{5.5}$$

Combining equations (5.4) and (5.5), the optimum on-resistance of switch *T2* can be calculated as follows:

$$r_{on} = \frac{1}{4\,\omega_{GBW}\,C_H} \tag{5.6}$$

Similar considerations hold for switch *T3* which also causes an additional pole together with the input capacitance of the operational amplifier main input. Switch *T1*, on the other hand, is not critical and can thus be designed using minimum dimensions.

When the ladder is finally attached to the auto-zeroed regulated cascode structure, one additional problem occurs. The ladder exhibits an output capacitance that depends on the digital code applied to the ladder and can be large for certain codes, as was discussed in section 2.5.3. The maximum output capacitance of the ladder in this implementation is 22 pF. This capacitance together with the transconductance g_m of transistor *T4* in figure 5.5 forms another pole in addition to the ones described in equation (5.3). The values of the transconductance as well as the on-resistance of switch *T2* that would be necessary to obtain real

1 The settling is slightly faster if the two poles are complex conjugate and spaced closely to each other. To have enough margin for process variations, however, it is best to choose the two poles real and identical, as indicated.

poles would require too large transistor dimensions to be implemented. A separation of the ladder, e. g. using switches, is not possible. Any resistance between the ladder output and the current-summing node of the regulated cascode would cause a voltage drop across that resistance due to the output current of the ladder. This voltage drop depends on the digital code applied to the ladder and would thus have the same consequences as an offset voltage between the output terminals of the ladder. Therefore, the output capacitance of the ladder has to be reduced in the auto-zero cycle. During normal operation the output capacitance of the ladder presents no problem, because the voltage at node V_{out} (see figure 5.5) is fixed.

A possible solution is to perform auto-zeroing of the two sides that generate I_{out} and I_{dump} (see figure 5.4) in two different clock cycles and to apply a digital code to the ladder that minimizes the output capacitance during the auto-zero phase. The digital code 0000 completely turns off the output current of the ladder, while code 1111 maximizes the output current, therefore the dump current is minimum. The clocking scheme of this operation as well as the codes applied to the ladder are depicted in figure 5.6.

5.2.3 Operational Amplifier Specifications

The auto-zero scheme as presented above puts more stringent constraints on the specifications of the operational amplifier than the regulated cascode alone would. The two most important parameters of an operational amplifier are the DC gain and the unity-gain bandwidth.

5.2.3.1 Gain Requirements

Three different necessary requirements have to be fulfilled by the regulated cascode structure that are influenced by the DC gain of the operational amplifier. First, the finite gain of the amplifier causes a deviation of the voltage V_{out} from V_{REF} in figure 5.5, as already discussed in sec-

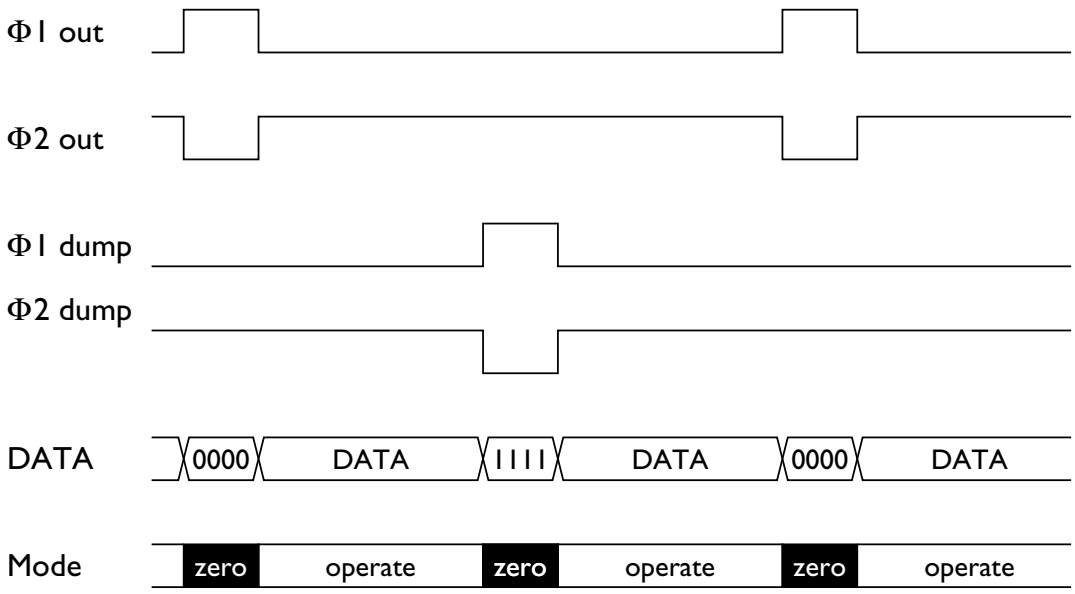

Figure 5.6: Clocking scheme of the D/A converter. The auto-zero phase is performed every other clock cycle to be able to minimize the output capacitance of the ladder by selecting a proper digital code.

tion 3.3.1. The minimum required gain is given in equation (3.3) and is repeated here with the values of figure 5.5 inserted:

$$A_{\min} = \frac{V_{\mathrm{T4}} + \sqrt{\dfrac{2I_{\mathrm{BIAS}}}{\mu_{\mathrm{eff}} C'_{\mathrm{ox}} \frac{W_4}{L_4}}} + V_{\mathrm{REF}}}{\Delta V_{\max}} \tag{5.7}$$

where $\Delta V_{\max}$ denotes the maximum allowed voltage error.

The auto-zero circuit, however, not only stores the offset of the amplifier, but also any non-zero input voltage that is necessary to generate the output voltage of the operational amplifier. Therefore, only the output changes due to the change in the saturation voltage of transistor *T4* for different currents have to be taken into consideration. Equation (5.7)

thus can be reduced to

$$A_{\min} = \frac{\sqrt{\frac{2I_{\mathrm{BIAS}}}{\mu_{\mathrm{eff}} C'_{\mathrm{ox}} \frac{W_4}{L_4}}}}{\Delta V_{\max}} \tag{5.8}$$

For the implementation described here, the minimum DC gain of the operational amplifier according to equation (5.8) is 70 dB as opposed to 89 dB predicted by equation (5.7).

The second requirement, namely low input impedance, has been discussed in detail in section 3.2. In the implementation of the successive-approximation A/D converter of chapter 3 the input impedance requirements were already fulfilled by choosing the operational amplifier gain as outlined above. For this converter, however, the requirements are more stringent due to the higher accuracy of the D/A converter. Therefore, a closer examination is necessary.

The current error caused by current division between the output impedance of the ladder and the input impedance of the regulated cascode structure is smaller that 0.5 LSB if the following inequality holds:

$$r_{\mathrm{in}} \leq \frac{R_{\mathrm{out}}}{2 \cdot 2^n} \tag{5.9}$$

where r_{in} is the small-signal input impedance of the regulated cascode structure, R_{out} denotes the output impedance of the MOSFET-only ladder for a specific digital code applied to the ladder, and n is the number of bits of the overall accuracy of the D/A converter, not only the stage resolution.

The input impedance of the regulated cascode structure depicted in figure 5.5 is given by

$$r_{\mathrm{in}} = \frac{1}{g_{\mathrm{m4}}(A+1)} \tag{5.10}$$

where A is the DC gain of the operational amplifier. Equation (5.10)

can be expanded to

$$r_{\text{in}} = \frac{1}{\sqrt{2I_{\text{D4}}\mu_{\text{eff}}C'_{\text{ox}}\dfrac{W_4}{L_4}\,(A+1)}} \tag{5.11}$$

with I_{D4} being the current flowing through transistor *T4*. This current depends on the ladder output current and thus on the digital code applied to the ladder. Therefore not only R_{out} but also r_{in} in equation (5.9) depend on the digital ladder code. For a given reference current and given dimensions of transistor *T4* the required minimum gain can be calculated for each digital code. The maximum gain requirement occurs at the codes where the current through transistor *T4* is minimum, i. e. code 1111 for I_{out} and code 0000 for I_{dump}. Although it may seem to be an overkill to calculate the required gain for each code and then find the maximum, simply taking the worst case for this calculation (which occurs at the minimum ladder output impedance and the minimum current through *T4*) would be too pessimistic. The resulting minimum gain using the exact procedure yields 68 dB and is almost the same as the requirement for the input voltage given above.

The last gain specification is imposed by the auto-zero cycle. Equation (5.2) shows that the original offset voltage V_{off} is attenuated by $(1 + A_{\text{aux}})$, the gain from the auxiliary input to the output. This attenuation must be large enough such that the sum of the remaining fraction of the original offset and the contribution due to the charge injection is smaller than the maximum voltage error that can be tolerated. Allowing the error due to the offset voltage to be half of the maximum error, the minimum gain of the auxiliary input is 53 dB, and thus the minimum DC gain of the main amplifier is 71 dB.

5.2.3.2 Speed Requirements

The necessary unity-gain bandwidth of the main operational amplifier in the operation phase is determined by the dynamic behavior

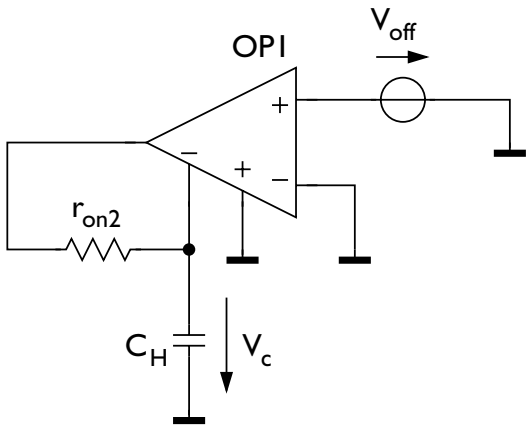

Figure 5.7: Small-signal equivalent circuit of the auto-zero phase to calculate the step response

of the input impedance. The calculation is identical to that outlined in section 3.3.2, and the result obtained in equation (3.9) also applies here.

As mentioned above, the amplifier from the auxiliary input to the output has a reduced unity-gain bandwidth compared to the main amplifier. Due to this unity-gain bandwidth reduction, the settling time during the sample phase of the auto-zero is the limiting factor of speed in this implementation, not the settling of the input impedance.

The settling behavior in the sample phase can be calculated using the small-signal equivalent circuit depicted in figure 5.7. r_{on2} denotes the on-resistance of switch *T2*. The cascode transistor in this phase only serves as a level shifter. Its pole is much higher than the dominant pole of the amplifier and the pole caused by the switch, so it can be neglected for this examination. The level shifting is also not important in small-signal calculations and is therefore omitted.

The dynamic behavior of the circuit can be determined by applying a voltage step V_{off} to the main input of the amplifier and calculating the response V_c across the hold capacitor C_H. It is assumed that switch

T2 of figure 5.5 is designed according to equation (5.6) such that equation (5.5) holds. The transfer function therefore consists of two identical real poles:

$$V_c = V_{off} \frac{A_{main}}{A_{aux}} \frac{1}{\left(1 + \frac{s}{2\,\omega_{GBW}}\right)^2} \tag{5.12}$$

where it has been assumed that the auxiliary gain is much larger than unity. ω_{GBW} is the reduced gain-bandwidth product of the amplifier using the auxiliary input. The step response can be calculated from equation (5.12) and is given by

$$V_c(t) = V_{off} \frac{A_{main}}{A_{aux}} \left[1 - (1 + 2\,\omega_{GBW} t)\, e^{-2\omega_{GBW} t}\right] \tag{5.13}$$

This equation serves as the basis to calculate the necessary unity-gain bandwidth of the auxiliary part of the amplifier. In this implementation, a unity-gain bandwidth of 55 MHz has been chosen to enable fast settling in the auto-zero phase. The unity-gain bandwidth of the main amplifier therefore equals 440 MHz.

5.2.4 Amplifier Architecture

To obtain unity-gain bandwidths in the range of 440 MHz, a single-stage architecture is advantageous over a two-stage topology due to the lower number of non-dominant poles. Moreover, short channel lengths have to be used for all transistors except for current sources. Although a high unity-gain bandwidth can be achieved this way, the gain that can be attained is small.

The gain can be increased by using regulated cascode structures for the operational amplifier [46]. The DC gain of the resulting operational amplifier is the product of the gain of the main amplifier on the one hand and the auxiliary amplifiers on the other hand. Because the gain can be increased to an (arbitrarily) high value by means of the regulated

cascodes, the main amplifier can be optimized for speed. In this way, DC gain requirements and speed considerations can be treated separately [86].

The simplified schematic of the operational amplifier for the current extraction circuit is depicted in figure 5.8. The main amplifier consists of transistors *T1*...*T10*. It has a folded cascode structure to obtain high bandwidth. Due to the use of minimum channel lengths for *T1*, *T2* and *T5*...*T12* for bandwidth reasons its gain is only 50 dB. The gain of the entire operational amplifier is then boosted to 110 dB by the regulated cascodes formed with the help of *OP1*...*OP4*.

The auxiliary input pair consists of transistors *T11* and *T12*. The smaller gain ($A_{\text{main}}/A_{\text{aux}} = 8$) is accomplished by reducing the bias current as well as the W/L ratio by a factor of eight compared to the main input pair *T1* and *T2*.

The auxiliary amplifiers *OP1*...*OP4* also utilize a folded cascode topology. These amplifiers have a gain of 60 dB and a unity-gain bandwidth of 300 MHz. When using a regulated cascode structure for gain enhancement, a pole-zero doublet is created around the unity-gain frequency of the auxiliary amplifier [46]. A pole-zero doublet causes a low-frequency settling component and thus is able to degrade the settling behavior of an operational amplifier [87]. To minimize the effect of this pole-zero pair, the unity-gain bandwidth of the auxiliary amplifier can be made large enough such that the doublet occurs at a frequency beyond the closed-loop frequency of the system. In many applications this is not necessary, however. The magnitude of the slow settling component caused by the doublet is proportional to the spacing between the pole and the zero of the doublet [87]. Therefore, it is sufficient to keep this spacing small. The zero caused by the regulated cascode is found at the unity-gain bandwidth of the auxiliary amplifier, $\omega_{\text{GBW}_{\text{aux}}}$. The pole is located at

$$p = -\frac{\omega_{\text{GBW}_{\text{aux}}}}{2}\left(1 + \sqrt{1 - \frac{4\,\omega_{\text{d}}}{A_{\text{aux}}\,\omega_{\text{GBW}_{\text{aux}}}}}\right) \tag{5.14}$$

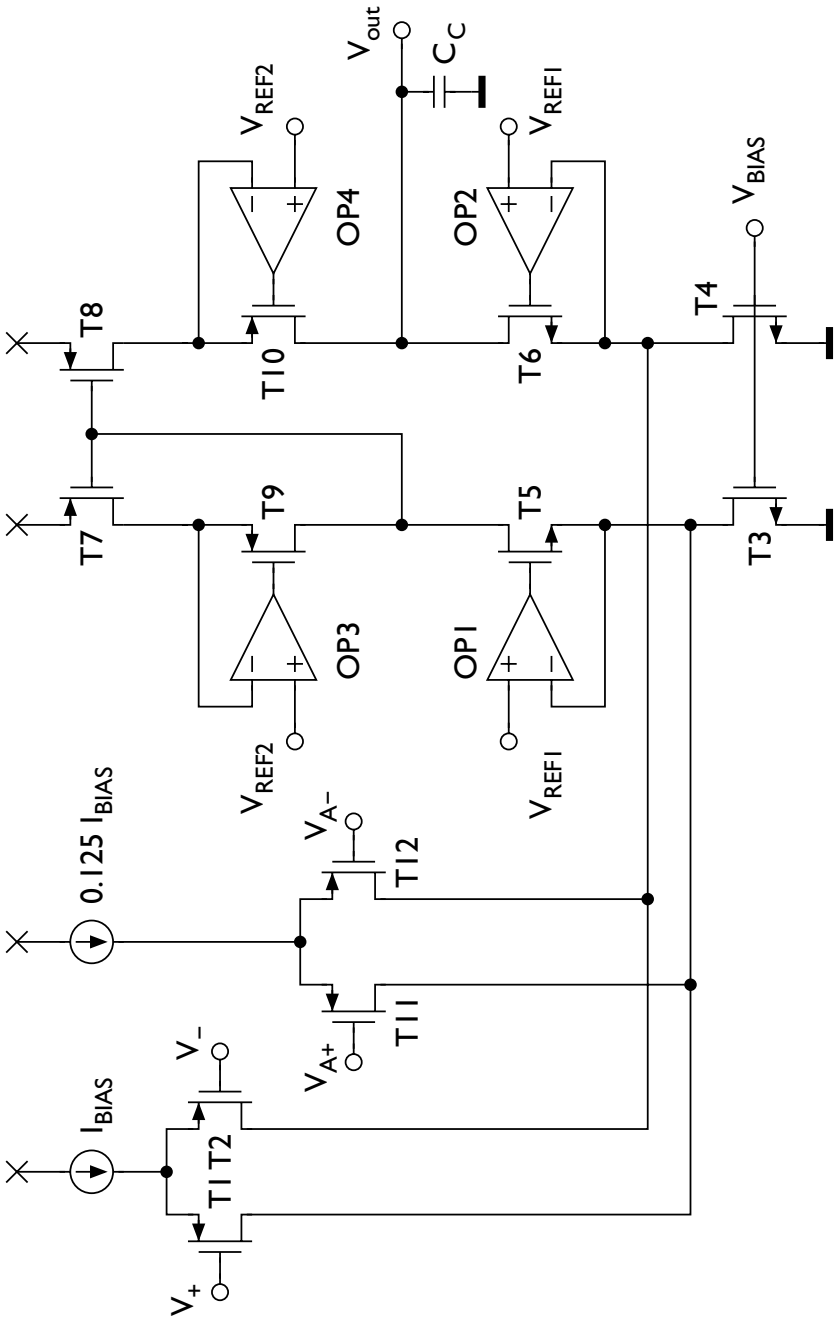

Figure 5.8: Schematic of the regulated cascode operational amplifier for the current extraction circuit

where ω_d denotes the dominant pole of the main amplifier without gain enhancement. Equation (5.14) shows that selecting the unity-gain bandwidth of the auxiliary amplifier sufficiently larger than the dominant pole of the main amplifier or choosing a comparatively large value for the auxiliary gain A_{aux} guarantees a close spacing between the pole and zero and thus minimizes the effect of the doublet.

The operational amplifier with the auxiliary amplifiers designed as outlined above shows no degradation in settling. Simulations yield a unity-gain bandwidth of 450 MHz with a phase margin of 68 degrees. The power consumption including the auxiliary amplifiers is 1.28 mA.

5.2.5 Reference Current Generation

The last task that has to be performed in the D/A converter is the generation of the reference current for the MOSFET-only ladder. This has to be accomplished internally by means of a resistor that is matched to the feedback resistors of the interstage gain block (see figure 5.1) to avoid the dependence of the D/A converter output voltage on the absolute value of an integrated resistor, as was pointed out in section 5.1.

The bias circuit is depicted in figure 5.9. Resistor R is matched to the feedback resistors as mentioned above. The reference current I_{REF} is given by

$$I_{REF} = \frac{V_{REFA} - V_{REFB}}{R} \tag{5.15}$$

The two reference voltages V_{REFA} and V_{REFB} are supplied externally.

The accuracy requirements for the reference current can be determined as follows. Any deviation of the reference current from its ideal value changes the output range of the D/A converter. This is equivalent to a D/A converter with ideal reference current plus a subsequent amplifier having a gain different from unity, as depicted in figure 5.10(a). To see the effects of the altered output range of the D/A converter, the additional gain stage can be relocated as shown in figure 5.10(b). Most im-

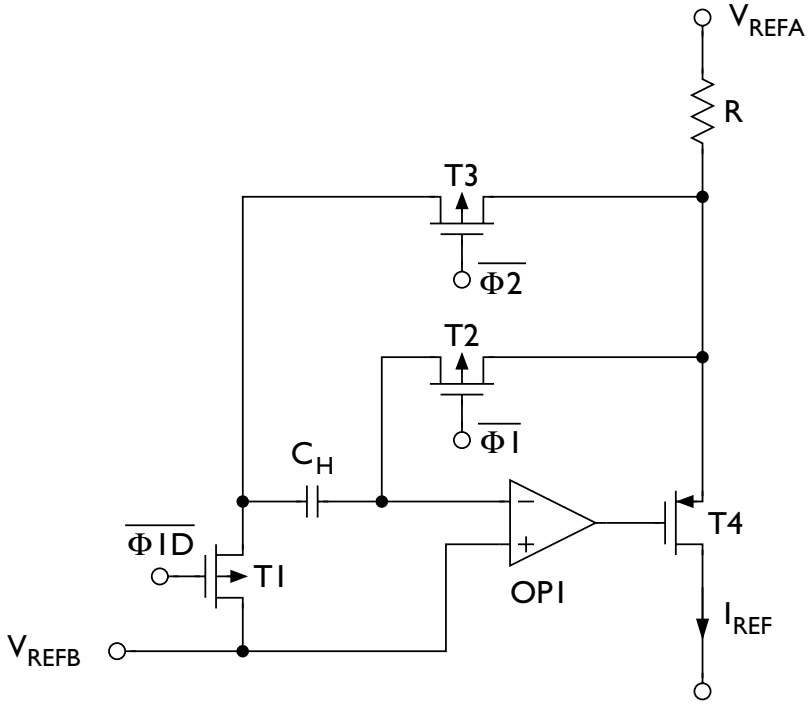

Figure 5.9: Internal reference current generation for the D/A converter

portant, the inaccurate reference current changes the conversion range of the entire stage, as indicated by amplifier ❶ in figure 5.10(b), as well as the output range (see amplifier ❸). Moreover, an additional error is generated at the input of the A/D subconverter which is represented by amplifier ❷.

If the reference current is inaccurate, but identical for all stages, the input and output errors cancel except for the input error of the first stage, so only the errors in the A/D subconverter paths remain. In other words, a change in the reference current for all stages only alters the conversion range of the entire A/D converter and causes a small error in the A/D subconverter path. As long as the deviation of the reference current from the ideal value is small, the error in the A/D subconverter path is irrelevant: as described in section 4.3.1, the accuracy of the A/D subconverter only has to match the resolution of one stage due to

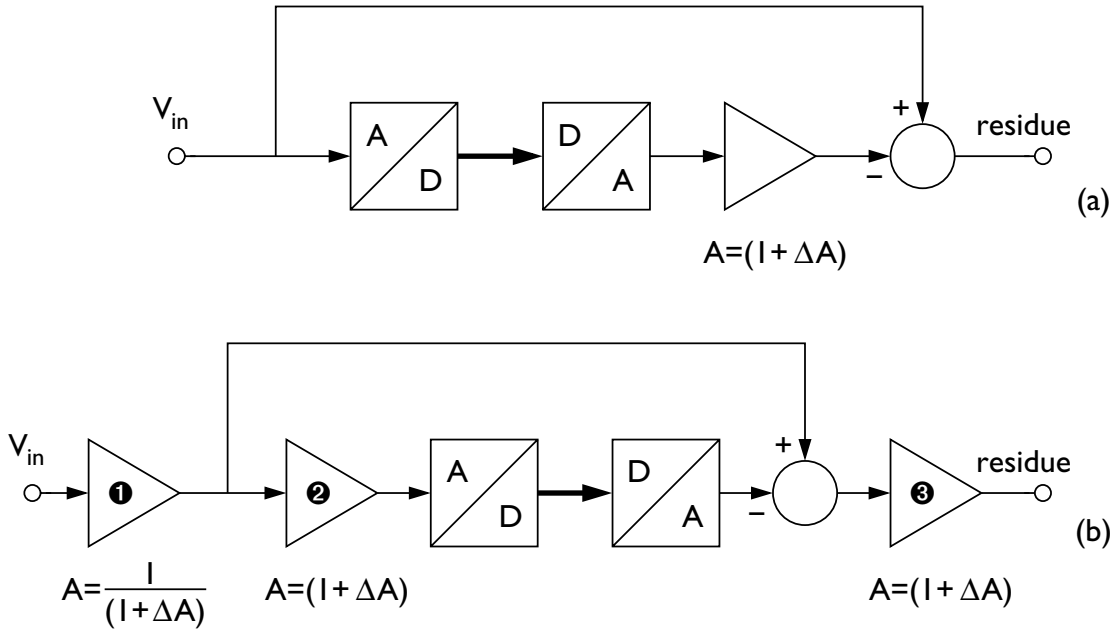

Figure 5.10: Change in output range of the D/A converter due to reference current inaccuracy (a), and the gain error moved towards the stage output (b). The sample-and-hold stage is omitted.

digital error correction. Therefore, any errors caused by an inaccurate reference current can be corrected as long as these errors are smaller than 0.5 LSB of the stage resolution.

If the reference currents differ for individual stages, however, the situation changes. For small deviations of the reference current from its ideal value, the error in the A/D subconverter path still can be neglected. The gain of amplifier ❸, however, is not cancelled by the input gain error of the next stage anymore and thus causes a gain error between the stages. This gain error cannot be corrected by digital error correction as discussed in section 4.3.3. Therefore, this error must be small enough such that it does not affect the linearity of the A/D converter.

The latter case applies to this D/A converter implementation because the reference current is generated locally in each stage to ensure matching of the resistors and thus its value is not equal for all stages. Therefore, equation (4.11) represents the upper limit of the gain error which then can be translated to a maximum allowed deviation of the reference current from its ideal value.

In this implementation, the maximum tolerable gain error ΔA equals $0.5 \cdot 2^{-9}$ for the first stage which yields a maximum current inaccuracy of $0.488\,\mu\text{A}$. The reference voltage $V_{\text{REFA}} - V_{\text{REFB}}$ has been chosen as $500\,\text{mV}$ which yields a value of $1\,\text{k}\Omega$ for resistor R in figure 5.9. Therefore, the error of the voltage across R must not exceed $0.488\,\text{mV}$. Two non-idealities prevent the voltage across R from being equal to $V_{\text{REFA}} - V_{\text{REFB}}$, namely the offset of the amplifier *OP1* and the finite loop gain. While the latter can be made as large as necessary by increasing the amplifier gain, the offset cannot be lowered to the required level without additional circuitry.

Therefore, auto-zeroing has to be applied to the operational amplifier. This has the additional benefit that not only the offset of the amplifier is stored, but also the input voltage necessary to generate the amplifier output due to finite gain. Thus, the DC gain of the operational amplifier only has to be sufficiently large to guarantee the accuracy of the offset cancellation. The full auto-zero circuit is shown in figure 5.9. As with the circuit used in the current extraction scheme (cf. figure 5.5), the loop remains closed during all clock phases to ensure stability. During clock phase $\Phi 1$ the offset and the input voltage of *OP1* are sampled onto the hold capacitor C_{H}. During $\Phi 2$ the capacitor is put into the feedback loop and cancels the offset. To minimize the residual offset, switches *T1* and *T2* are equipped with dummy switches. In addition, *T1* is switched by a delayed version of $\Phi 1$.

The reference current generation as well as the current extraction circuit described before exhibit high accuracy. Part of the necessary accuracy stems from the fact that the D/A converter is built around the

MOSFET-only ladder and thus has to satisfy the offset requirements of the ladder. But the stringent accuracy requirements also reflect the fact that the D/A converter is the most important building block of a pipeline stage in terms of linearity. Therefore, the D/A converter always is a building block that requires high-accuracy components, regardless of the specific implementation.

5.3 A/D Subconverter

In contrast to the D/A converter, the A/D subconverter has comparatively relaxed accuracy requirements. Due to the corrective feedback loop of the digital error correction it is sufficient if the A/D subconverter matches the stage resolution which is four bit in this pipeline. Therefore, the design can be focused on speed rather than accuracy. The natural choice of the A/D subconverter architecture is thus a full flash topology.

5.3.1 Fully Differential A/D Converters

The entire signal path in the pipeline is fully differential, therefore this also expands to the A/D subconverter. For an A/D converter fully differential operation means that the *difference* of the input voltages $V_{\text{in}+}$ and $V_{\text{in}-}$ is compared to the *difference* of the reference voltages $V_{\text{R}+}$ and $V_{\text{R}-}$ of each code. This also means that a positive as well as a negative value of the reference voltage of each code has to be generated by a resistor string. The basic principle is depicted in figure 5.11. For clarity, two separate resistor strings are drawn that generate the necessary positive and negative reference voltages. Comparators with two differential input pairs then compare the two voltage differences as discussed above.

The circuit can be simplified by generating all reference voltages by means of one single resistor string [88, 89]. The resulting converter is

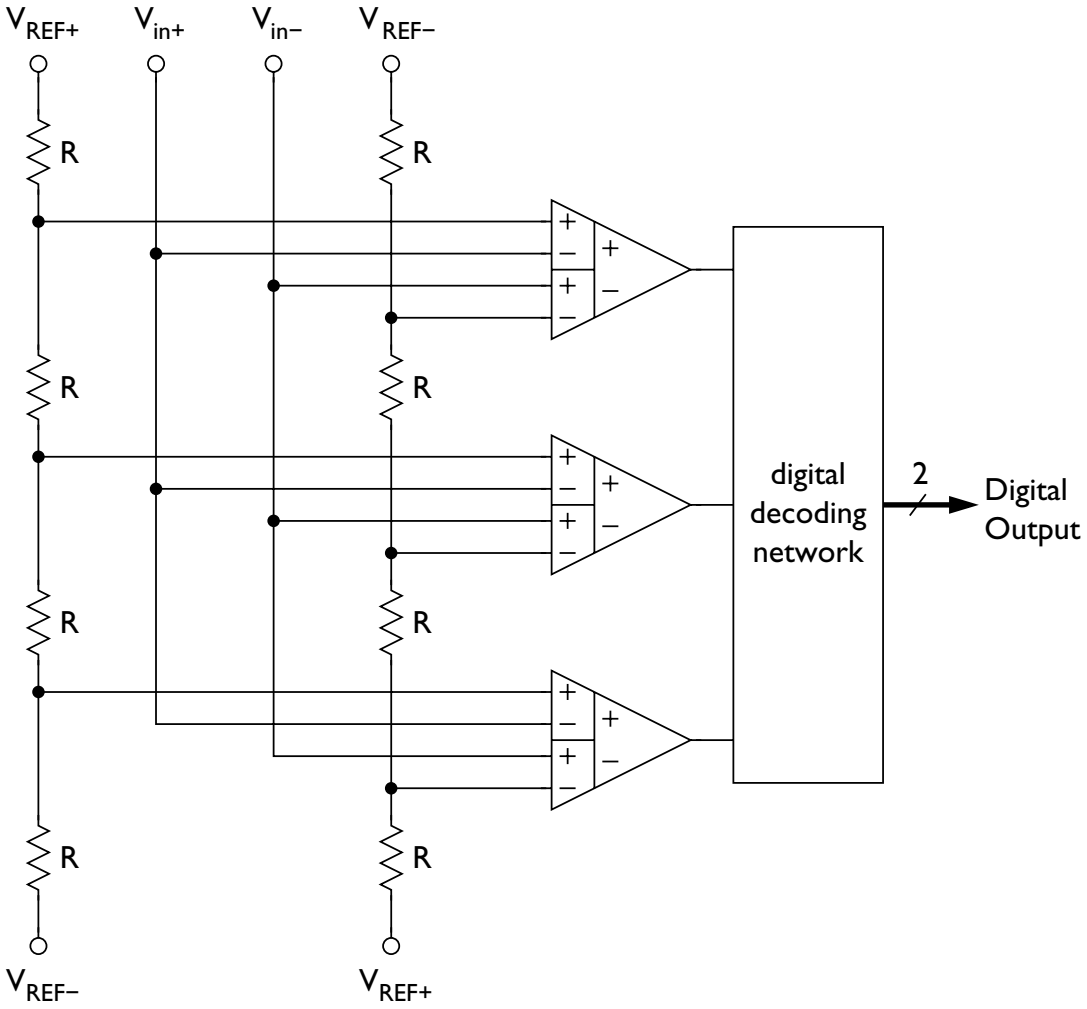

Figure 5.11: Basic principle of a fully differential 2 bit A/D converter

depicted in figure 5.12. In this implementation every tap of the resistor string is used twice to provide positive and negative values with respect to a common-mode voltage that is equal to the potential in the middle of the resistive string. The decision whether to use one or two reference strings will be made after the values of the unit resistors have been determined.

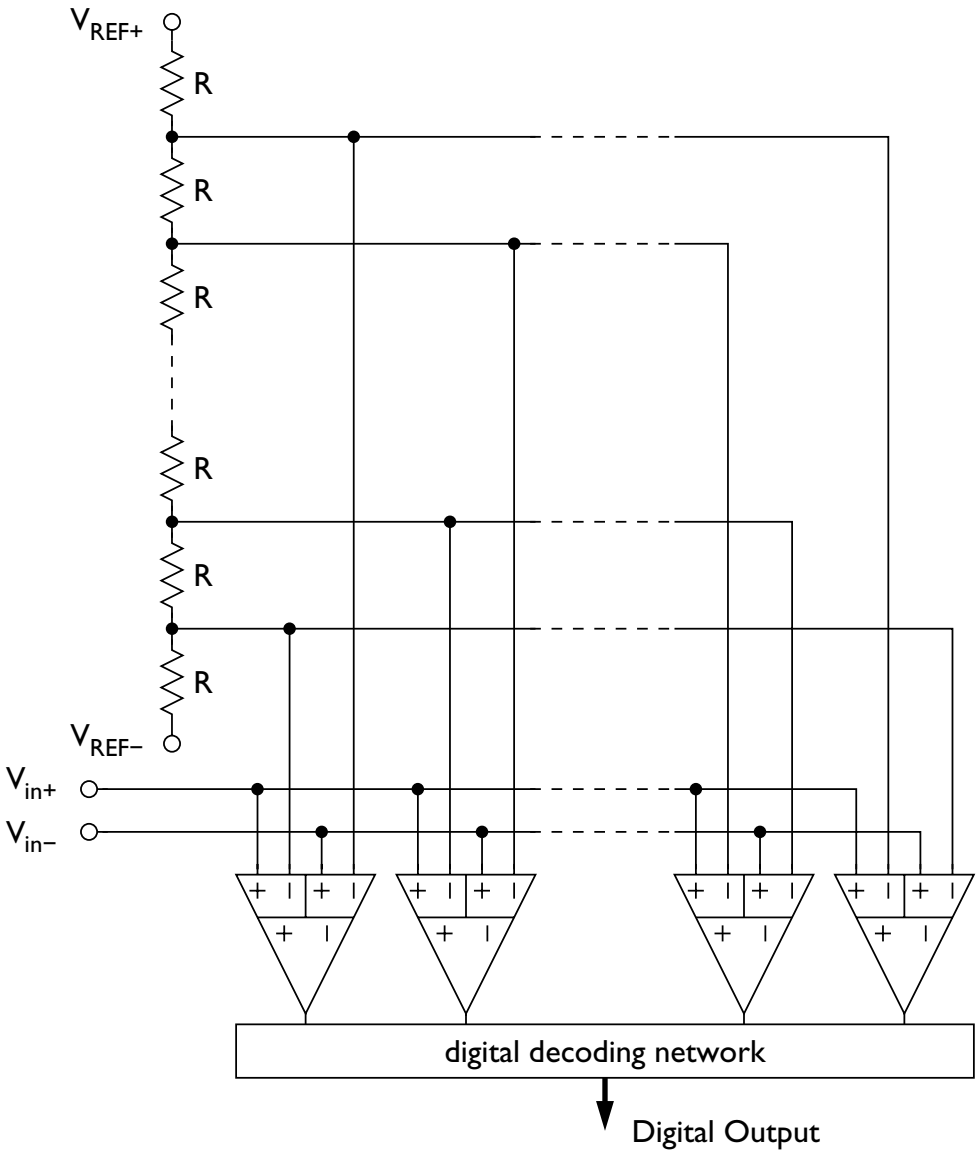

Figure 5.12: Fully differential A/D converter with a single resistor string

5.3.2 Reference Ladder Bowing

In principle, the resistors of the reference string should be made as high-ohmic as possible to minimize the power consumption of the A/D converter. A limit is set, however, by reference ladder bowing, which means a disturbance of the reference voltages that are generated by the resistor string due to interaction with the comparators connected to this string [66, 90].

Two types of interaction between the comparators and the reference string exist. DC reference ladder bowing is caused by loading of the resistive string due to input currents of the comparators. This type of error only plays a role in bipolar designs because CMOS converters exhibit no DC input current. More severe, however, is AC reference ladder bowing. This error is caused by capacitive coupling between the input signal and each reference voltage node. The coupling is caused by the gate-source capacitance of the input transistors as depicted in figure 5.13(a). The resulting equivalent circuit which shows the coupling of all comparator inputs is depicted in figure 5.13(b).

Different analytical models have been derived to calculate the error due to reference ladder bowing. One model takes into account different values of coupling capacitances for different comparators depending on the input voltage [90]. More important, however, is the understanding of the underlying mechanism. Reference ladder bowing is most severe for flash A/D converters that operate without a sample-and-hold circuit. In this case a constantly changing input voltage is present at the comparator inputs. When operating in a pipeline stage, however, the A/D subconverter input is isolated from the input signal by the stage sample-and-hold circuit. Nevertheless, reference ladder bowing influences the settling time and must thus be taken into consideration.

By examining the A/D converter with a single reference string in figure 5.12 it can be seen that the two differential input pairs of the comparators that share one tap of the resistive string are connected to complementary input voltages. Therefore their influence on the ref-

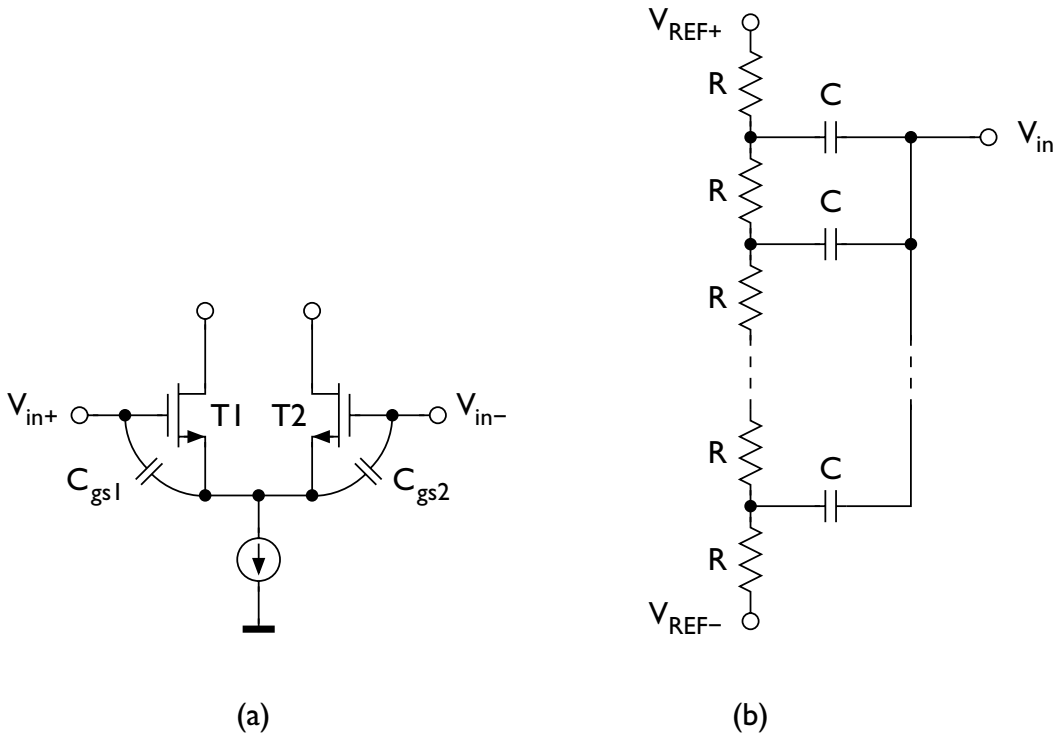

Figure 5.13: Mechanism of the coupling between the input signal and the reference voltages causing A/D converter reference ladder bowing (a), and the resulting equivalent circuit (b).

erence string due to capacitive coupling nearly cancels. The solution with a single reference string is thus preferable if no other architectural constraints force the utilization of two separate strings.

In this implementation a single resistive string has been chosen. Due to the cancellation of the capacitive coupling a unit resistance of 418 Ω is sufficient to keep the reference ladder bowing small. The total resistance of the entire resistive string is thus 6685 Ω, which yields a comparably small current of 150 μA with a reference voltage of 1 V. The layout consists of a single polysilicon strip with a width of 2 μm. The reference voltages of the individual codes are created by small taps to this string.

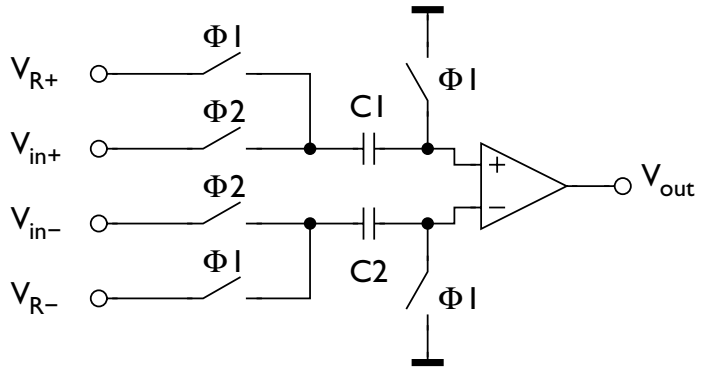

Figure 5.14: Basic fully differential comparator with switched-capacitor input

5.3.3 Fully Differential Comparator

The fully differential operation of the A/D subconverter also requires comparators that are able to compare two differential input signals, as discussed in section 5.3.1. The commonly used fast comparator structures, however, are only capable of comparing two voltages, but not two voltage differences. Therefore, the differencing operation has to be carried out by a special circuit before the result is fed into the actual comparator.

A straightforward solution is using a switched-capacitor circuit to generate the voltage differences [6, 91]. The principle is depicted in figure 5.14. During clock phase Φ1 the capacitors are charged via the reference voltages. In clock phase Φ2 the difference between the input signal and the reference voltages is generated. This implementation, however, is sensitive to differences in the common-mode voltages of the input signals and the reference signals, respectively. The common-mode rejection can be increased by modifying the circuit, but at the expense of using four capacitors instead of two [92]. In a technology where only metal sandwich capacitors are available, and therefore even small capacitors consume a considerable amount of silicon area, a different solution is preferable.

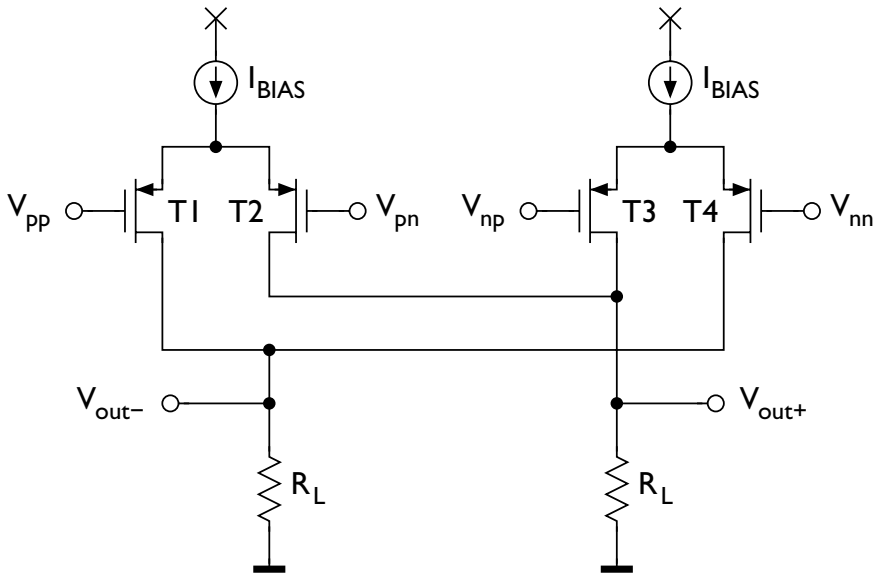

Figure 5.15: Differential-difference amplifier for the comparator of the A/D subconverter

The difference of two differential input signals can also be formed with a differential-difference amplifier (DDA) [93]. It basically consists of two interleaved differential pairs. The DDA implemented for this comparator is depicted in figure 5.15. The output voltage is given by

$$V_{\text{out+}} - V_{\text{out-}} = A\left[(V_{\text{pp}} - V_{\text{pn}}) - (V_{\text{np}} - V_{\text{nn}})\right] \qquad (5.16)$$

The gain A of the DDA equals $g_{\text{m}} \cdot R_{\text{L}}$ where it is assumed that all transistors in figure 5.15 have the same transconductance g_{m}. A polysilicon resistor has been chosen as a load for the DDA to maximize speed. The value of the gain is not critical. A moderate gain, however, reduces the offset voltage of the subsequent comparator.

The voltage comparator following the DDA is of the same type as the comparator used for the successive-approximation A/D converter which is described in detail in section 3.4. Due to the low common-

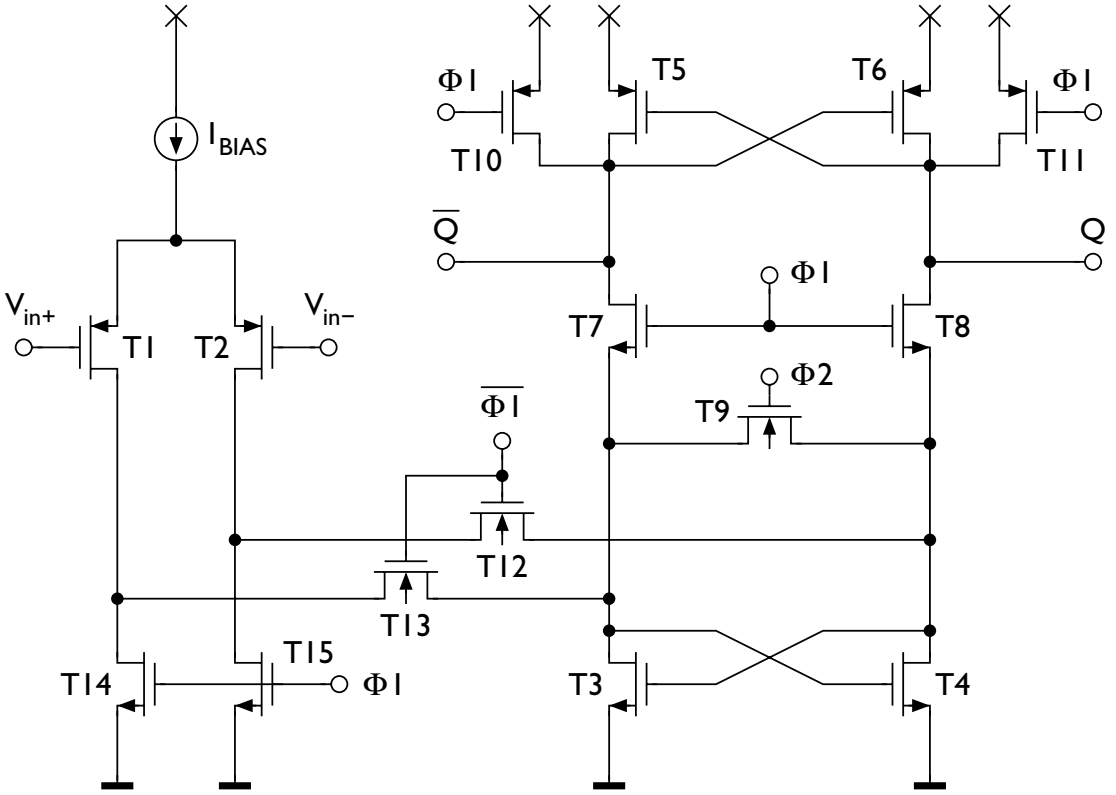

Figure 5.16: Voltage comparator for the A/D subconverter

mode output voltage of the DDA, however, the input differential pair of the comparator has to be designed with PMOS transistors as opposed to the NMOS implementation for the successive-approximation A/D converter. The schematic of the resulting comparator is depicted in figure 5.16. The output signal of the DDA is amplified by the differential pair consisting of transistors *T1* and *T2*. Regeneration of the NMOS latch (*T3* and *T4*) starts with the falling edge of clock phase Φ2. The rising edge of clock phase Φ1 then activates the entire CMOS latch which generates the full-swing output.

The low common-mode voltage at the input of the comparator, however, causes one additional problem. If e. g. the output Q of the

comparator is high after the regeneration has taken place, the voltage at the (drawn) drain side of transistor *T2* in figure 5.16 is much higher than on the source side. Therefore, drain and source of *T2* change places which yields a current flowing through the input differential pair. This current is in the order of two times the bias current I_{BIAS} of the input stage and flows during the entire clock phase $\Phi 1$. The power consumption of the comparator is thus increased substantially, which cannot be neglected due to the high number of comparators (60 units in 4 flash A/D subconverters) used.

To prevent this unnecessary additional current from flowing, the input differential pair has to be isolated from the CMOS latch in clock phase $\Phi 1$. This is accomplished by the switches *T12* and *T13* in figure 5.16. The additional transistors *T14* and *T15* provide a path for the bias currents of *T1* and *T2* when the differential pair is not connected to its load. The clock signals for transistors *T12*...*T15* are derived from clock signal $\Phi 1$ locally to ensure the proper switching sequence.

The resulting fully differential comparator including the DDA frontend consumes 51 µA of current when operated with a 5 MHz clock. The typical (simulated) propagation delay from the falling edge of $\Phi 2$ to stable outputs is 1.75 ns.

The comparators together with the polysilicon reference string described in the last section form two different full flash A/D subconverters for the pipeline stages. The decision levels of the A/D subconverters in the first three stages are shifted towards more positive values by 0.5 LSB to be able to perform addition-only digital error correction, as described in section 4.3.1. The A/D subconverter for the last stage, however, does not need any shift of the decision levels, which would only cause a shift of the transfer function of the entire pipelined A/D converter. Each subconverter occupies a silicon area of 0.065 mm^2 including the reference string, all comparators, thermometer code detection and the output decoder.

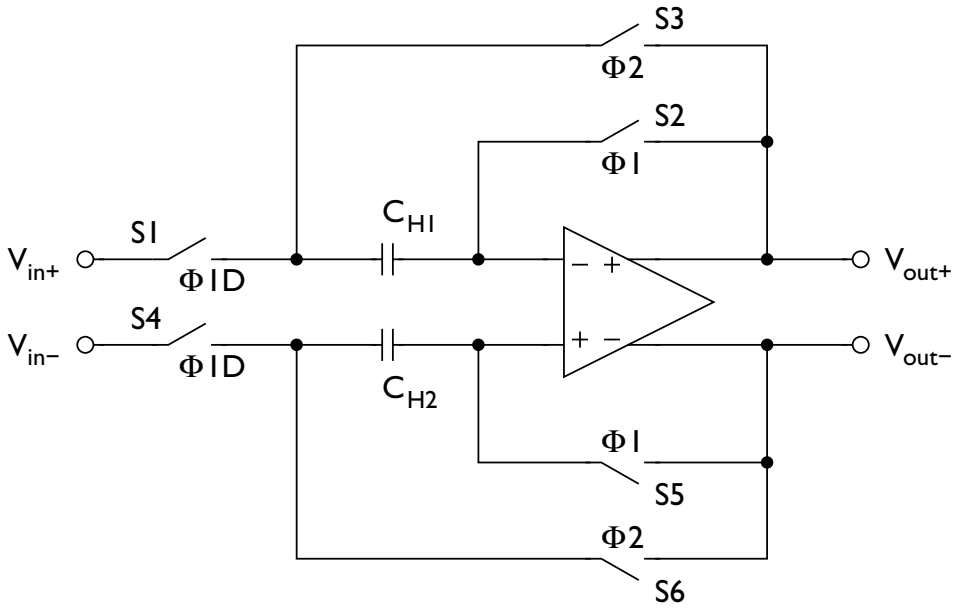

Figure 5.17: Architecture of the input and stage sample-and-hold circuit

5.4 Sample-and-Hold Circuit

After the discussion of the two subsystems that actually perform the analog-to-digital conversion, namely the A/D subconverter and the D/A converter that enables multistage conversion, another building block remains that is equally important in pipelined A/D converters. The sample-and-hold (S/H) circuit at the input of the A/D converter as well as the interstage S/H circuits not only make the pipelined operation possible, but also influence the dynamic performance of the entire A/D converter.

5.4.1 General Architecture

The basic architecture of the sample-and-hold circuit used in this implementation is depicted in figure 5.17 [39, 70]. The architecture can be kept simple because, unlike in most other pipeline implementations,

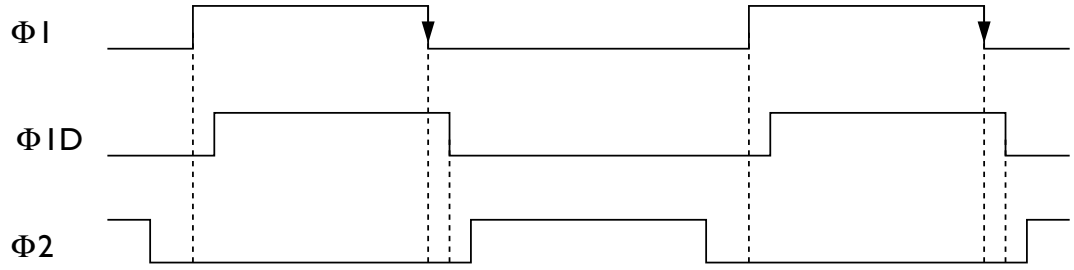

Figure 5.18: Timing of the sample-and-hold circuit. The delayed clock is provided to minimize signal-dependent charge injection.

the sample-and-hold circuit does not have to provide any gain, as was discussed in section 5.1. The two sample-and-hold circuits that are used at the input of the A/D converter and between the stages have the same topology. The sampling capacitance, however, is scaled due to different noise requirements. Consequently, the operational amplifier in the stage S/H circuit is also a scaled version of the amplifier used at the input.

The basic operation of the sample-and-hold circuit is as follows. During clock phase Φ1 the unity-gain feedback around the operational amplifier, which is provided by switches *S2* and *S5*, establishes virtual ground at the input nodes of the amplifier. The input signal is sampled on the two hold capacitors C_{H1} and C_{H2}, respectively, via the two sampling switches *S1* and *S4*. At the end of the sample phase, the feedback switches *S2* and *S5* open first, while the sampling switches *S1* and *S4* are driven by a delayed version of clock Φ1. The detailed timing of the S/H circuit is shown in figure 5.18. This switching sequence reduces the signal-dependent charge injection [94]. The opening of the feedback switches *S2* and *S5* causes charge injection on the hold capacitors. The input nodes of the amplifier, however, are always at the same common-mode potential, independent of the input signal. The charge injection is therefore also independent of the input signal. Moreover, if the two switch transistors and the hold capacitors are perfectly matched,

the pedestal on the hold capacitors due to charge injection only causes a change in the input common-mode voltage. Even in the presence of mismatch the charge injection only causes offset, but no distortion. When finally the sampling switches *S1* and *S4* open, the charge released by these switches flows into the signal source, because no conducting path from the hold capacitors to a low-ohmic node exists any more. Therefore, the charge injection caused by *S1* and *S4*, which is signal-dependent, does not cause any error on the hold capacitors.

During clock phase $\Phi 2$ the hold capacitors are connected to the output of the operational amplifier. Therefore the sampled and held input voltage is restored at the amplifier output. Due to the fact that the outputs of the amplifier always return to the common-mode voltage during the sample phase (caused by the unity-gain feedback), the maximum output voltage step the operational amplifier has to perform is only half of the full scale voltage.

The choice of the hold capacitor value is a trade-off between noise requirements on the one hand and speed and power consumption on the other hand. The sampling action adds kT/C noise to the system which only can be reduced by increasing the hold capacitance C_H. A large capacitance, on the other hand, increases the load of the operational amplifier and thus decreases the unity-gain bandwidth for a given power consumption. Moreover, the area needed for large capacitance values in a technology that only provides metal sandwich capacitors quickly grows very large. The selected value of the hold capacitors therefore is always a compromise. For the input sample-and-hold circuit hold capacitors with a value of 5 pF have been chosen. The reduction of the signal-to-noise ratio (SNR) due to kT/C noise is below 1.3 dB for this capacitor value, the silicon area for both input hold capacitors, however, is already $0.065\,\mathrm{mm}^2$.

Due to the gain between the individual pipeline stages the hold capacitor values can be reduced in the subsequent stages. Although all stages could have smaller hold capacitors than their predecessors, the reduction has been carried out only once to reduce the number of dif-

ferent stages that have to be designed. Although the interstage amplifier between the first and the second stage has a gain of eight, only a factor of five in capacitor reduction has been chosen in order to minimize the influence of the kT/C noise of the subsequent stages on the overall signal-to-noise ratio.

5.4.2 Switch Design

An important step in the design of a sample-and-hold circuit after having established the architecture and the value of the hold capacitance is the design of the switches. The type and the dimensions of the switches not only determine the charge injection that can cause non-linearity and offset, but also the bandwidth and the stability of the sample-and-hold circuit is affected. The switches can be divided into two sets which correspond to the sample and the hold mode, respectively.

5.4.2.1 Hold Mode

The switches *S3* and *S6* in figure 5.17 provide feedback during hold mode. They are not involved in the more sensitive sampling process, and thus their requirements are somewhat relaxed. The situation during the hold mode can best be viewed by means of a simplified schematic, as depicted in figure 5.19. r_{on3} denotes the on-resistance of switch *S3* (or *S6*). Three capacitances have to be considered: the hold capacitance C_H, the input capacitance C_{in} of the operational amplifier and the bottom-plate capacitance C_{BP} of the hold capacitor. Due to the usage of metal sandwich capacitors the achievable capacitance per area is small which leads to substantial values of the bottom-plate capacitance because the area occupied by the hold capacitors is large. Using all metal layers, the bottom-plate capacitance is 15% of the hold capacitance. Omitting the lowest metal layer would decrease the bottom-plate capacitance but at the cost of even higher consumption of silicon area.

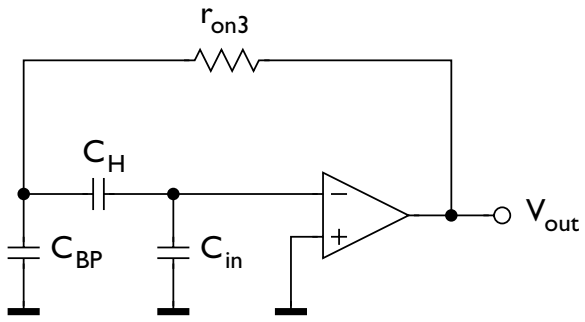

Figure 5.19: Schematic of the S/H circuit during hold mode. For simplicity, only a single-ended version is drawn.

The effective capacitance is given by

$$C_{\text{eff}} = C_{\text{BP}} + \frac{C_{\text{H}}\, C_{\text{in}}}{C_{\text{H}} + C_{\text{in}}} \tag{5.17}$$

This capacitance together with the on-resistance of *S3* forms a pole in addition to the dominant pole of the operational amplifier. The situation is identical as in the discussion of the auto-zero circuit in section 5.2.2. The location of the resulting poles is thus also given by equation (5.3). The only difference is the open-loop position of the pole caused by switch *S3* which now is located at

$$\omega_{\text{s}} = \frac{1}{r_{\text{on3}}\, C_{\text{eff}}} \tag{5.18}$$

To obtain two identical real poles, the equality given in equation (5.5) must hold, which is repeated here for convenience:

$$\omega_{\text{s}} = 4\,\omega_{\text{GBW}} \tag{5.19}$$

The necessary on-resistance of switches *S3* and *S6* can thus be calculated as

$$r_{\text{on3}} = \frac{1}{4\,\omega_{\text{GBW}}\, C_{\text{eff}}} \tag{5.20}$$

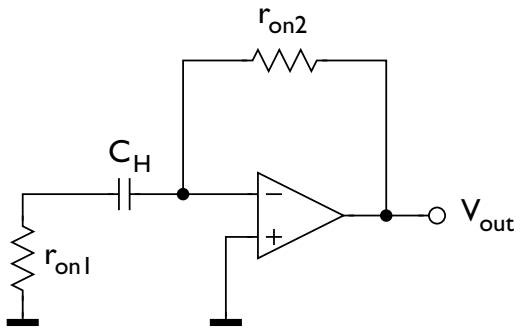

Figure 5.20: Simplified schematic of the S/H circuit during sample mode

The unity-gain bandwidth of the operational amplifier is 136 MHz (see section 5.4.3). Together with an effective capacitance C_{eff} of 825 fF equation (5.20) yields an on-resistance of 350 Ω, a value which is easy to implement.

No special measures are necessary to suppress charge injection. Transmission gates have to be used, however, to keep the changes of the on-resistance at an acceptable level over the full output voltage swing.

5.4.2.2 Sample Mode

During sample mode, four switches are closed: *S2* and *S5* provide the virtual ground nodes at the input of the operational amplifier. The input signal is connected to the hold capacitors via the switches *S1* and *S4*. As in the hold mode, a simplified single-ended schematic can be used to determine the dynamic behavior. The circuit is depicted in figure 5.20. Both the bottom-plate capacitance as well as the input capacitance of the operational amplifier can be omitted in this case. On the other hand, the on-resistances of both switches *S1* and *S2* have to be taken into consideration. Due to the second resistance in the loop, the transfer function and thus the location of the poles differs from the results obtained above. The calculation of the poles is lengthy

but straightforward and yields

$$p_{1,2} = -\frac{1}{2}\left(\omega_s + \frac{r_{on1}}{r_{on1} + r_{on2}}\,\omega_{GBW}\right) \cdot \left(1 \pm \sqrt{1 - \frac{4\,\omega_s\omega_{GBW}}{\left(\omega_s + \frac{r_{on1}}{r_{on1} + r_{on2}}\,\omega_{GBW}\right)^2}}\right) \tag{5.21}$$

where ω_s is defined as

$$\omega_s = \frac{1}{C_H(r_{on1} + r_{on2})} \tag{5.22}$$

In contrast to the situation in the hold mode, there is one more degree of freedom in this case: the on-resistances of the switches *S1* and *S2* can be set independently. The optimum is reached if the on-resistance of *S1* approaches zero, which essentially would lead to the same situation as in the hold mode. A more realistic choice from the implementation point of view, however, is to choose both on-resistances being equal.

To obtain two identical poles, assuming that the on-resistances r_{on1} and r_{on2} are equal, as suggested above, the following equality must hold:

$$\omega_s = \left(1.5 + \sqrt{2}\right)\omega_{GBW} \approx 3\,\omega_{GBW} \tag{5.23}$$

Although this seems to be a relaxed requirement compared to equation (5.19) in the hold mode, the on-resistance of the switches in the sample mode must be much smaller due to the larger effective capacitance. Using the value of the unity-gain bandwidth given above, equation (5.23) yields an on-resistance of 39 Ω for each of the switches *S1* and *S2*.

Dummy switches are used for the transistors forming the switches *S2* and *S5* to reduce the pedestal on the hold capacitors due to charge injection. Although it has been stated in section 5.4.1 that charge injection

at these nodes only alters the common-mode voltage of the input signal or causes offset in case mismatch occurs, it is good practice to keep the pedestal small and thus reduce the deviation from the common-mode voltage. Special attention also has to be paid to the input switches *S1* and *S4*. In contrast to *S2* and *S5*, these switches are exposed to the full input voltage swing. Therefore, the on-resistance changes substantially for varying input signals. The variation can be reduced by using a transmission gate that is designed for symmetry such that the on-resistance is equal for maximum and minimum input voltage.

5.4.3 Fully Differential Operational Amplifier

The design of the switches is closely linked to the unity-gain bandwidth of the operational amplifier, as equations (5.19) and (5.23) demonstrate. Choosing the amplifier bandwidth too high demands unnecessarily low on-resistance of the switches and thus large transistor dimensions which in turn increases the charge injection. Moreover, the wide-band noise caused by the amplifier increases. Therefore, the smallest possible unity-gain bandwidth that satisfies the settling requirements has to be determined.

Once again, a distinction between the sample mode and the hold mode has to be made. In hold mode, the settling of the output voltage determines the speed. When the switches *S3* and *S6* in figure 5.17 are designed according to equation (5.20), the following step response of the single-ended circuit shown in figure 5.19 results:

$$V_{\mathrm{out}}(t) = V_{\mathrm{c}}\left[1 - (1 + \omega_{\mathrm{GBW}}\, t)\, \mathrm{e}^{-2\omega_{\mathrm{GBW}} t}\right] \tag{5.24}$$

V_{c} denotes the sampled and held input voltage across the hold capacitor C_{H}. The output voltage has to fulfill two settling requirements. The first building block in the pipeline stage that processes the output of the sample-and-hold circuit is the A/D subconverter. By the time the converter makes its decision, the output signal of the sample-and-hold

circuit must have settled to four bit accuracy. The settling to the full 13 bit accuracy can take place simultaneously with the settling of the D/A converter and is this thus less critical.

In the sample mode, the speed is dictated by the required acquisition time. For a sample-and-hold stage, the acquisition time is defined as the time difference between the sample command and the actual tracking of the input signal. For the sample-and-hold circuit of figure 5.17 this is the time the operational amplifier needs to re-establish virtual ground at its input nodes after switching from hold mode to sample mode. Using the single-ended model in figure 5.20, the output waveform after the rising clock edge of clock phase $\Phi 1$ can be calculated as

$$V_{\text{out}}(t) = V_{\text{out0}} \left[1 + \omega_{\text{GBW}} \left(1 + \frac{\sqrt{2}}{2} \right) t \right] \mathrm{e}^{-\omega_{\text{GBW}} \left(1 + \frac{\sqrt{2}}{2} \right) t} \qquad (5.25)$$

where V_{out0} denotes the voltage at the output of the operational amplifier during hold mode. In this implementation, the bandwidth of the operational amplifier is determined by the required acquisition time of the sample-and-hold circuit and not by the hold mode settling requirements. The necessary unity-gain bandwidth is 130 MHz.

The gain requirements are straightforward. Any gain error of the input sample-and-hold stage due to finite gain of the operational amplifier only causes a change in the conversion range of the entire A/D converter, but does not affect linearity. Gain errors between the stages, however, must be avoided, as discussed in section 4.3.3. A consultation of the block diagram of one pipeline stage (figure 5.1) shows that two building blocks can cause an interstage gain error: the sample-and-hold circuit at the input of each stage and the gain stage at the output. The maximum gain error that can be tolerated can be split between the two blocks in different ways. Allowing the same amount of error in both blocks puts stringent gain requirements on the gain stage amplifier due to the gain of eight while completely relaxing the gain specifications for the interstage sample-and-hold amplifier. A better solution is to design

both amplifiers for the same DC gain. This yields a minimum required gain of 80 dB for both operational amplifiers.

The choice of the operational amplifier architecture is influenced by the load at its output. The capacitive load in the sample mode is mainly caused by the hold capacitors. A small additional contribution stems from the input capacitance of the 15 comparators in the A/D subconverter. The total load capacitance in the sample mode is thus 5.8 pF. During the hold phase the load drops to 1.5 pF. Although the smaller load capacitance is advantageous from a speed point of view, problems can arise regarding stability if the load capacitance is also used as a compensation capacitance to shift the dominant pole of the operational amplifier towards lower frequencies, as is the case with most single-stage amplifier architectures. Under such circumstances the lowering of the load capacitance causes a decrease of the phase margin and thus the stability.

In addition to the load capacitance, a resistive load of 2 kΩ is present at the output stemming from the resistive feedback of the interstage gain block (cf. figure 5.1). To drive a resistive load while maintaining high gain, an output stage is necessary. Another benefit of this additional stage is the decoupling of the unity-gain bandwidth and the load capacitance. If the capacitance decreases in this case, only the output pole is shifted towards higher frequencies, which increases stability rather than decreases it. The obvious solution would be using a source follower to drive the load resistance. Due to the low supply voltage, however, a transistor in common-source configuration would limit the output signal swing too much. This is a common problem in low-voltage implementations. An inverting output stage in class-AB operation can be used to maximize the output voltage swing [95, 96]. In this implementation, however, the large load capacitance makes a high transconductance of the output transistor necessary to push the output pole to a sufficiently high frequency. The bias current that is required to establish this transconductance is much higher than the maximum current flowing through the resistive load. Therefore the power saved

by class-AB operation is negligible such that the complexity added by this configuration cannot be justified.

The operational amplifier is thus equipped with an inverting second stage in class-A operation using Miller compensation. The input stage consists of a folded cascode structure using the regulated cascode gain-boosting technique. The simplified schematic of the operational amplifier is depicted in figure 5.21. Due to the fully differential output a common-mode feedback is required. The implemented sample-and-hold architecture requires the operational amplifier to be active in both clock phases (cf. figure 5.17). Therefore, only a continuous-time common-mode feedback can be used. The schematic of the feedback circuit is depicted in figure 5.22. The passive network consisting of *R1*, *R2*, C_{C1} and C_{C2} generates the common-mode voltage at the output. An error amplifier compares this voltage to the reference voltage V_{CM}. The output is then fed to the bias transistors *T3* and *T4* of the main operational amplifier. The error amplifier is kept simple and consists only of an input differential pair and diode-connected load transistors. The cascode is provided to ensure good matching between transistors *T20*, *T21* and *T3*, *T4*.

The resulting operational amplifier for the input sample-and-hold circuit has a simulated unity-gain bandwidth of 136 MHz for typical process values. The phase margin is 60 degrees in sample mode when a load capacitance of 5.8 pF is connected to the output. The DC gain equals 140 dB. The current consumption is 4.64 mA which yields 11.6 mW of power. Most of the power consumption is due to the output stage.

As discussed in section 5.4.1, the interstage sample-and-hold block is a scaled version of the input sample-and-hold circuit. Consequently, the operational amplifier is also scaled down due to the smaller capacitive load. While maintaining all other specifications, its power consumption could thus be reduced to 7.11 mW.

The entire sample-and-hold circuit is capable of a sampling rate up to 5 Msample/s. The silicon area is 0.15 mm^2 which is mainly due to the

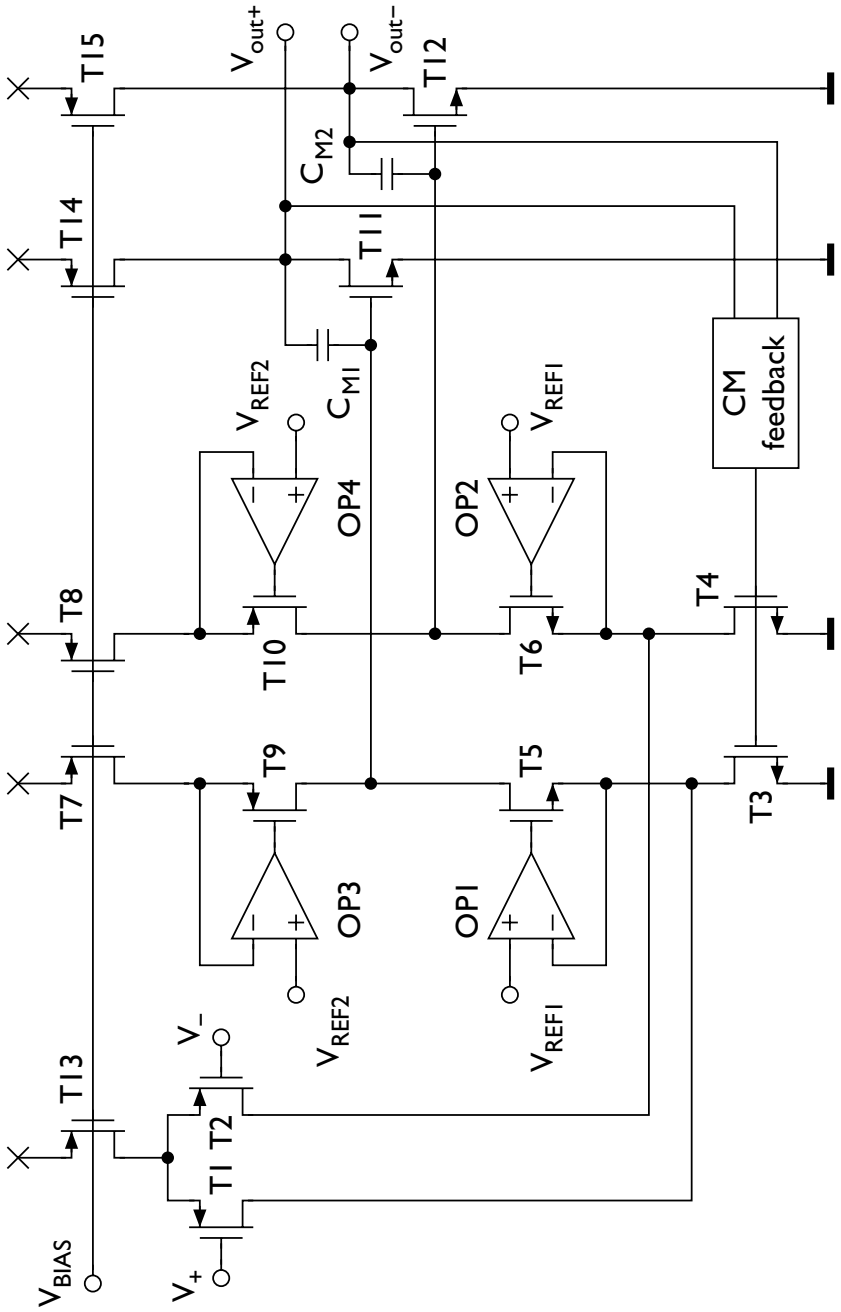

Figure 5.21: Simplified schematic of the two-stage Miller-compensated operational amplifier used in the sample-and-hold circuit

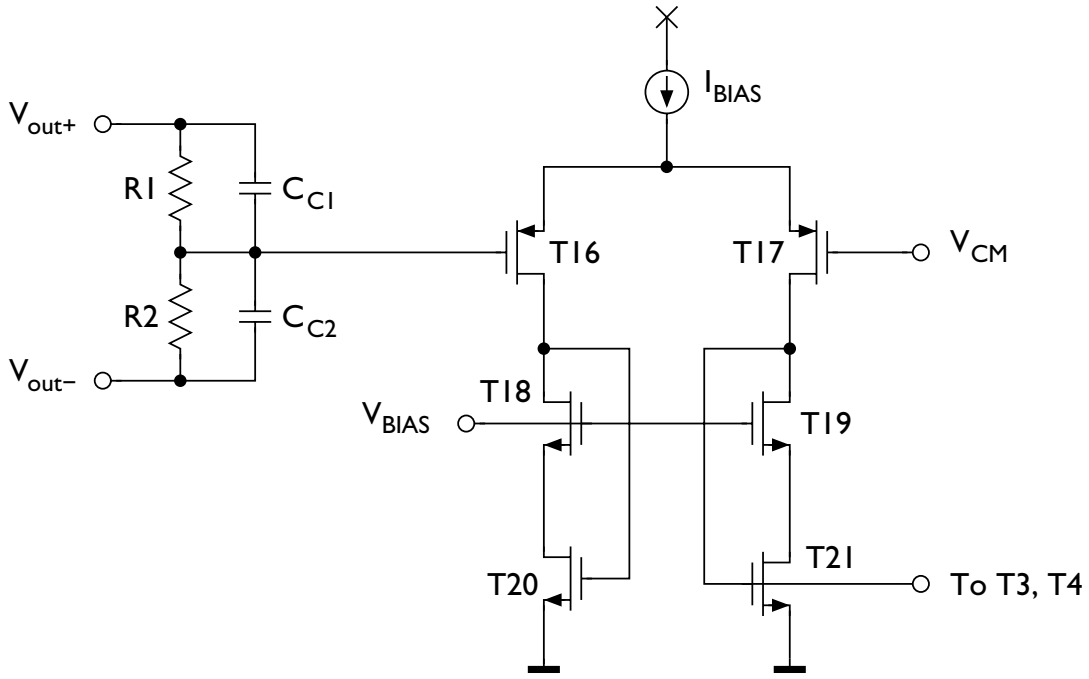

Figure 5.22: Continuous-time common-mode feedback circuit

area needed by the hold capacitors as well as the compensation capacitors of the operational amplifier. Special attention has to be paid to the layout. Any asymmetry of the switch layout causes the charge injection and clock feedthrough to be different for the two differential signal paths and thus generates offset and distortion. Therefore a highly symmetrical layout of the entire sample-and-hold circuit is necessary.

5.5 Interstage Gain

The last building block to be described, although less complex than the ones before, is very important for the functionality of the pipeline stage. The subtraction of the D/A converter output from the analog, sampled and held input signal is accomplished in the interstage gain block. The resulting difference signal is then amplified and output as the residue of the stage.

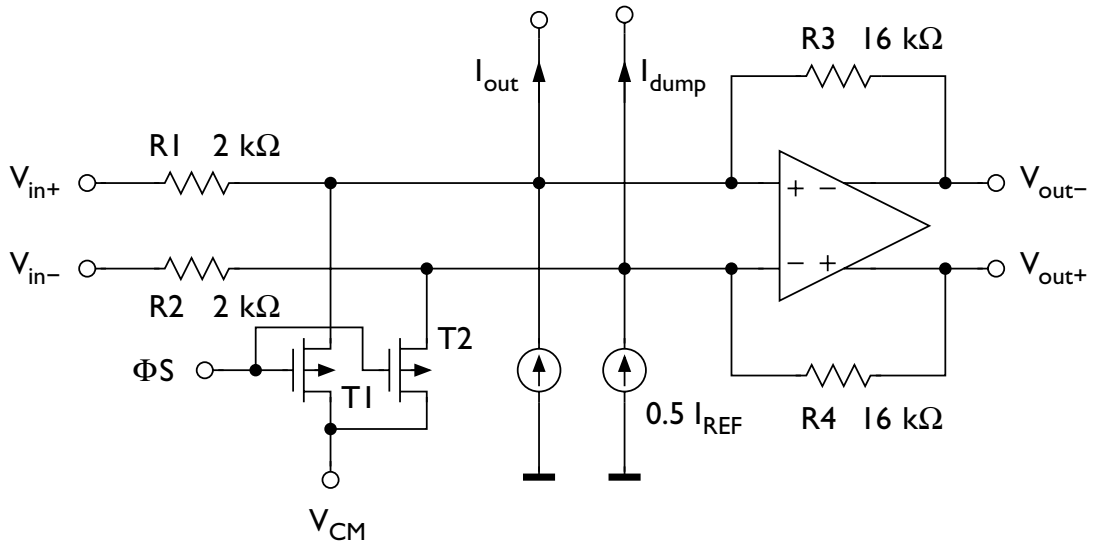

Figure 5.23: Schematic of the interstage gain block. The D/A converter output current is subtracted from the analog input, and the difference is amplified.

The schematic of the gain stage is depicted in figure 5.23. The reasons for choosing this architecture have already been discussed in section 5.1. So only the details follow here. The analog input signal stemming from the sample-and-hold circuit is converted to a current by means of the input resistors *R1* and *R2*, both of which have a value of 2 kΩ. This resistance results from matching the full scale current flowing through *R1* and *R2* to the reference current of the D/A converter.

The output currents of the D/A converter are fed into the input nodes of the operational amplifier. The two current sources in figure 5.23 convert the pseudo-differential output of the D/A converter to real differential signals.[2] The resulting currents that flow through

2 The two output currents of the D/A converter, namely I_{out} and I_{dump}, are not real differential signals. The dump current I_{dump} equals $I_{\text{REF}} - I_{\text{out}}$, but not $-I_{\text{out}}$, as it would be necessary for a differential system. Therefore, the author suggests the term *pseudo-differential* for these currents.

the feedback resistors *R3* and *R4* are thus the differences between the currents representing the analog input and the output currents of the D/A converter. These differences are converted back to the voltage domain by the feedback resistors. When only the analog input voltage is considered (omitting the D/A converter output for a moment), the entire block is a normal fully differential voltage amplifier with resistive feedback. The implemented gain of eight is dictated by to the stage resolution of four bit.

One specialty of the circuit are the switches *T1* and *T2* in figure 5.23. The gain stage amplifies the residue of the first stage which has a maximum amplitude of ± 1 LSB. The residue equals the difference between the input signal and the D/A converter output signal, as described above. When looking at the block diagram of one pipeline stage (cf. figure 5.1), however, it can be seen that the output of the D/A converter is only stable after the A/D subconverter has made its decision and the D/A converter has settled. The analog input signal, on the other hand, appears instantaneously at the input of the gain stage operational amplifier. Therefore, a large signal is applied to the input of the operational amplifier until the D/A converter output settles. This would cause the output stage of the operational amplifier to saturate, breaking up the feedback loop due to the inability of the amplifier to perform its regulation action. Recovery from this overload is slow, therefore the output of the amplifier has to be prevented from going into saturation.

The two switches *T1* and *T2* clamp the input of the operational amplifier to the input common-mode level during the first third of the clock phase. Until then the D/A converter output has settled enough such that the residue is within the range that can be amplified by the gain stage.

The specifications for the operational amplifier are straightforward. The gain requirements have been discussed already in section 5.4.3 in the context of the sample-and-hold amplifier. The speed requirements can be determined using the step response of the gain stage. When applying

a voltage step to the terminals $V_{\text{in}+}$ and $V_{\text{in}-}$ in figure 5.23, the output voltage equals

$$V_{\text{out}}(t) = -V_{\text{in}} \frac{A\,A_{\text{CL}}}{1 + A + A_{\text{CL}}} \left(1 - \mathrm{e}^{-\frac{\omega_{\text{GBW}}}{1+A_{\text{CL}}} t} \right) \tag{5.26}$$

where A is the DC gain of the amplifier, and A_{CL} denotes the ideal closed-loop gain of the feedback system which is eight in this implementation. The reduction of the effective bandwidth of the operational amplifier due to the gain is evident.

The type of output load the amplifier must be able to drive is similar to that of the sample-and-hold amplifier: the capacitive load is due to the hold capacitors of the subsequent stage, and the load resistance stems from the resistive feedback. Therefore, the same amplifier architecture as in the sample-and-hold stage can be used for this operational amplifier. The unity-gain bandwidth, however, is higher in this case due to the gain of eight that has to be implemented. The resulting amplifier has a unity-gain bandwidth of 235 MHz and a phase margin of 60 degrees when a load capacitance of 1.2 pF is applied to the output. The power consumption is 12.6 mW.

5.6 Assembly of the Pipeline Stages

The building blocks that have been described in the last four sections form one pipeline stage, which in turn is the main building block of the pipelined A/D converter. The first three pipeline stages are identical, except for the different sample-and-hold block of the first stage. The fourth stage, however, only consists of a sample-and-hold stage and the A/D subconverter. The D/A conversion and the generation of a residue is not necessary in the last stage. Moreover, the reference levels of the A/D subconverter in the last stage are not shifted, as discussed in section 5.3.

The raw digital codes output by the individual pipeline stages are fed to a digital block that includes the shift registers that are necessary to store the digital data of the individual samples concurrently processed in the pipeline and performs the addition-only error correction described in section 4.3.1. The resulting digital code representing the value of the analog input sample is then stored in an output register from where it is sent to the pads.

5.7 Experimental Results

The pipelined A/D converter has been implemented in a digital 0.25 µm CMOS process with six layers of metal and one layer of polysilicon. A high-ohmic polysilicon layer with a sheet resistance $R_\square$ of 220 $\Omega/\square$ is available to integrate resistors, but no special capacitors are provided. The most important drawback for analog designs, however, is the low power supply voltage of only 2.5 V. The consequences of the lack of compact capacitors and the low supply voltage on the architecture and implementation of the design have been discussed during the description of the individual building blocks.

The chip photograph is depicted in figure 5.24. The four pipeline stages are clearly visible. The regular field of ten by seven squares in the middle of the first (leftmost) stage are the two 5 pF hold capacitors of the input sample-and-hold circuit. Unfortunately, most of the other details are hidden by tiles. These small rectangles consisting of metal layers increase the uniformity of the metal coverage which is required by chemical mechanical polishing (CMP) used at the IMD (inter metal dielectric) planarisation step.

The A/D converter operates from a 2.5 V supply. Analog and digital supply as well as the supply for the output pad drivers have been separated to shield the analog part from power supply glitches caused by the digital circuitry. The power consumption of the entire A/D converter is 117.5 mW. The analog part consumes 115.25 mW. The digital circuitry

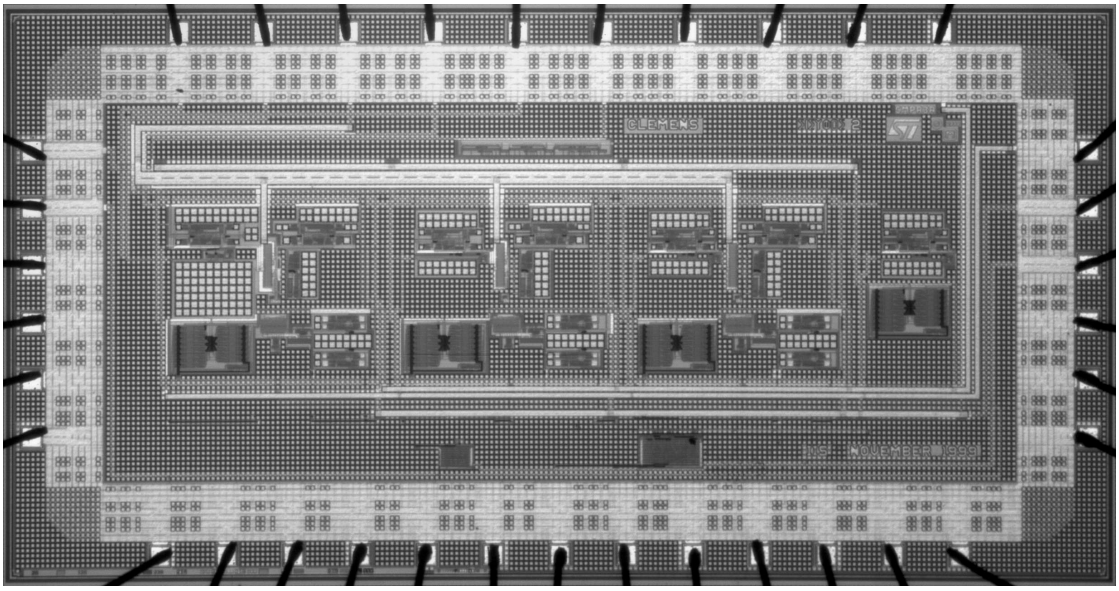

Figure 5.24: Chip photograph of the pipelined four-stage A/D converter

needs 2.25 mW of power, including the output buffers, at a sampling frequency of 5 MHz.

The output spectrum of an A/D converter is one of the most important measurements to judge the dynamic performance of the converter. The spectrum of the pipelined A/D converter at a sampling frequency of 5 MHz is depicted in figure 5.25. The input frequency of 2.09 MHz, which is almost the Nyquist frequency, has been chosen such that the aliased harmonics are widely spaced and clearly visible. The total harmonic distortion of the A/D converter is -70 dB measured at an input level of -1 dB with reference to full scale. The spurious-free dynamic range (SFDR) is 76 dB.

Additional important measurements of an A/D converter are the integral non-linearity (INL) and the differential non-linearity (DNL). Often, the INL and DNL of an A/D converter are specified using static measurements. It is more meaningful, however, to measure these non-linearities dynamically at the maximum conversion speed. Only then

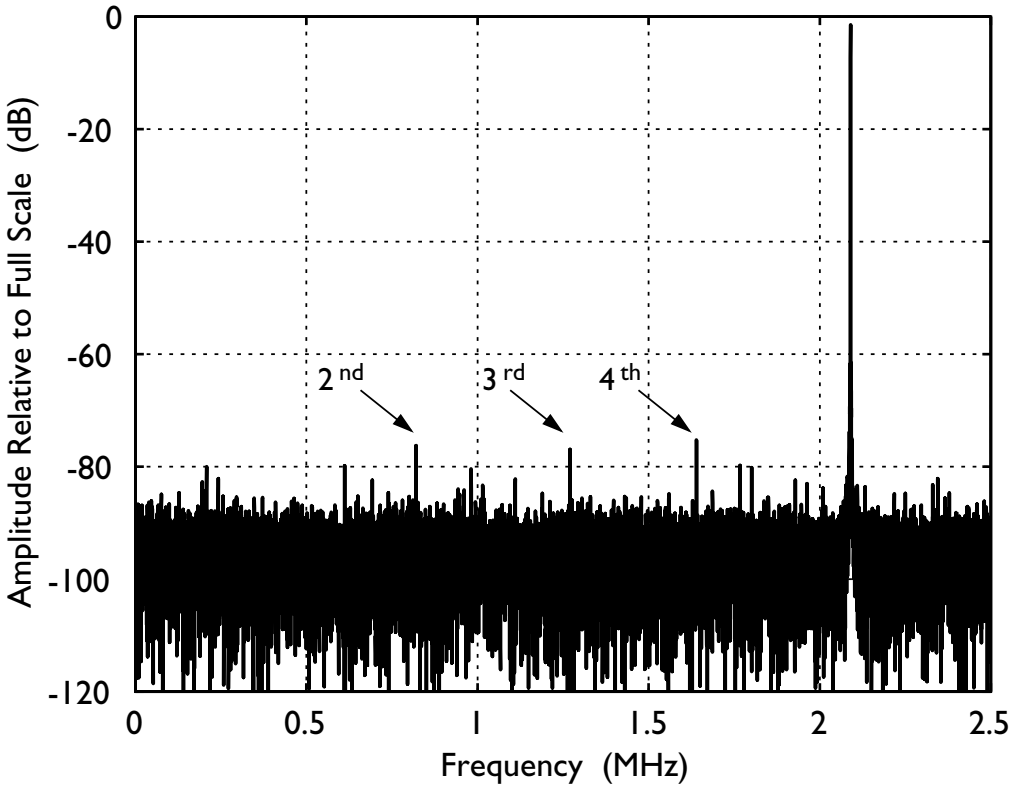

Figure 5.25: Output spectrum of the pipelined A/D converter. The input frequency is 2.09 MHz, the sampling frequency equals 5 MHz. The FFT was performed using 32768 data points and a Kaiser-Bessel window with $\alpha = 4$.

errors due to the limited bandwidth of the individual building blocks are noticeable, so that the true dynamic performance can be measured. INL and DNL are obtained by measuring a histogram of the output codes of the converter, from which the code density can be calculated. The measured INL and DNL at a conversion speed of 5 Msample/s is depicted in figure 5.26. 2 097 152 codes were used for this histogram test. The measured DNL is completely within the ± 1 LSB limit for 13 bit accuracy. The INL shows a maximum value of 4.4 LSB. The jumps noticeable in

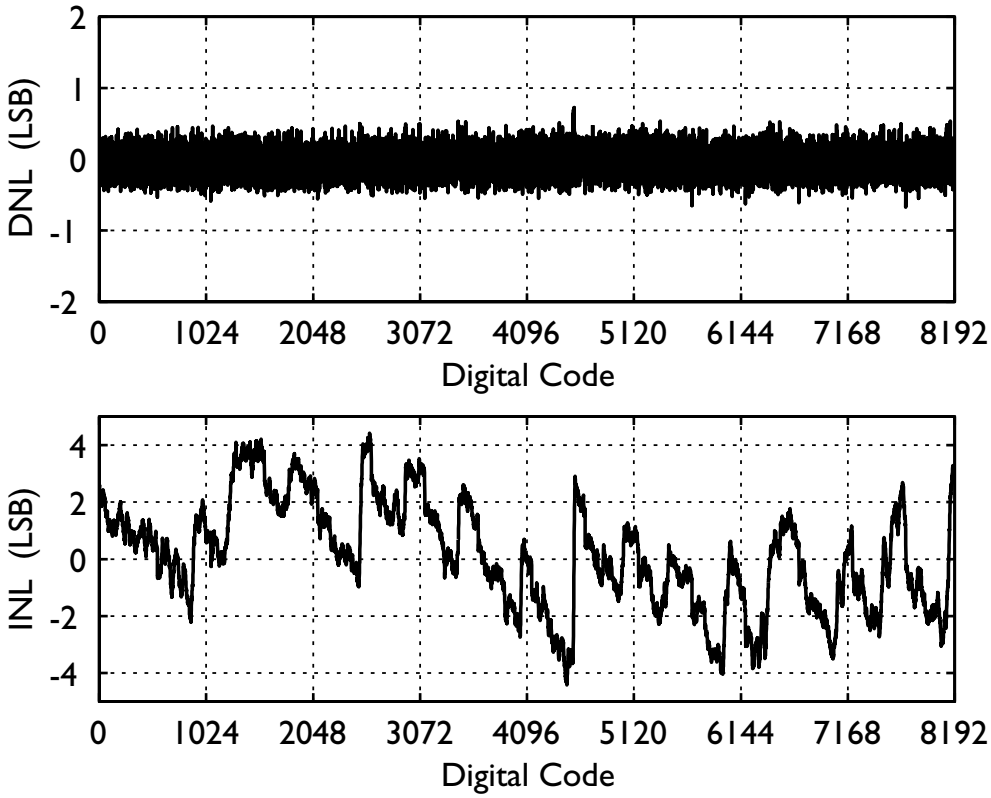

Figure 5.26: Measured INL and DNL at 5 Msample/s with an input frequency of 1.99 MHz

the INL measurement stem from mismatch in the MOSFET-only ladder circuits. The INL performance could therefore be improved by increasing the dimensions of the ladder transistors. This, however, would also decrease the conversion speed, as discussed in section 5.2.1.

To prove the dynamic performance of the input sample-and-hold circuit, the total harmonic distortion of the converter has been measured at a sampling frequency of 5 Msample/s with different input signal frequencies up to the Nyquist frequency. The results are shown in figure 5.27. No significant degradation can be seen up to the Nyquist frequency.

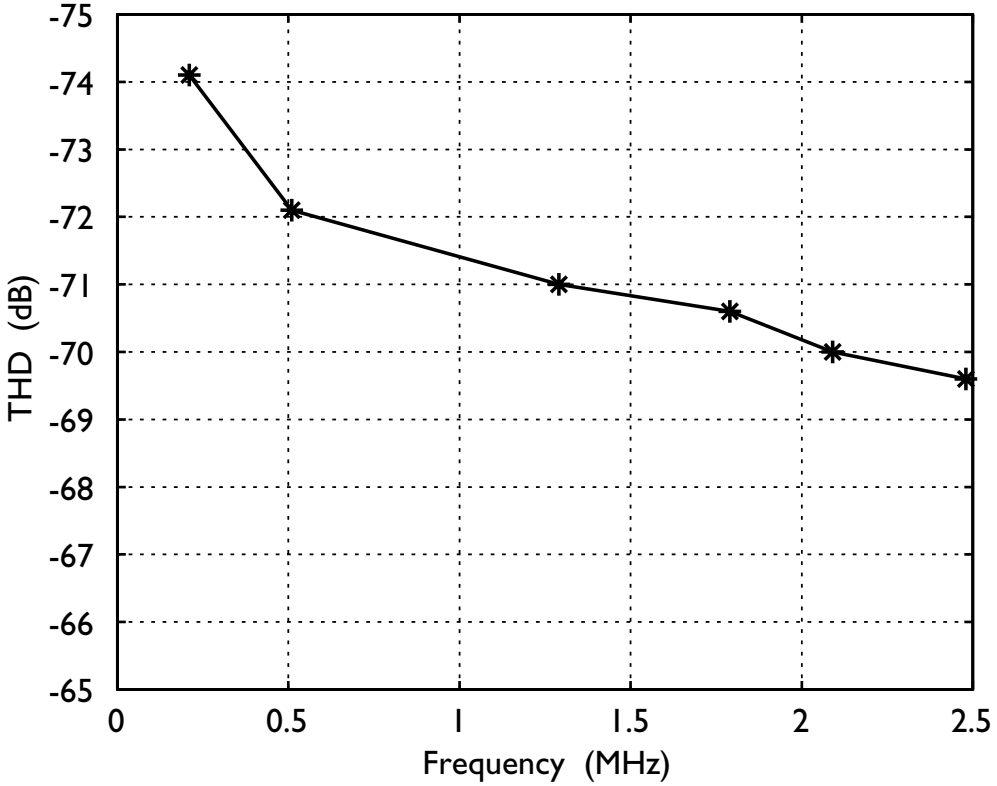

Figure 5.27: Measured THD versus input frequency at 5 Msample/s.
The measured input frequencies are denoted by stars.

A summary of the A/D converter characteristics and the measured performance is given in table 5.1.

5.8 Summary

A pipelined 13 bit A/D converter with four bits per stage is described in this chapter. The D/A converter, which is the most important building block in the pipeline stage with respect to linearity, is based on the MOSFET-only ladder described in chapter 2.

Table 5.1: Characteristics of the pipelined A/D converter

Process	0.25 µm Single-Poly 6-Metal CMOS
Chip area including pads	4364 µm × 2207 µm
Resolution	13 bit
Sampling frequency	5 Msample/s
Measured INL	4.4 LSB
Measured THD	−70 dB
Spurious-free dynamic range	76 dB
Input range	$2\,V_{pp}$
Power consumption	117.5 mW @ 2.5 V

While the successive-approximation A/D converter discussed in chapter 3 proved that the MOSFET-only ladder can be used in A/D converters with more than ten bit resolution, this design is aimed to test the MOSFET-only ladder at higher speeds.

Several changes are necessary to adapt the standard architecture of a pipeline stage to the needs of the MOSFET-only ladder. The most important point in this context is the offset sensitivity of the ladder which has to be taken into account. If this is done, however, the ladder yields a compact, accurate and versatile D/A converter.

A second design challenge is the low power supply voltage that is accompanied by today's deep submicron technologies. Several analog design techniques cannot be used any more due to the lack of headroom, others become more complex.

The experimental results of the A/D converter show that the MOSFET-only ladder can indeed be used at higher sampling frequencies.

Good total harmonic distortion was achieved at a conversion speed of 5 Msample/s. The measured dynamic DNL is at a 13 bit level, while the measured INL shows the trade-off between accuracy, which can be increased by better matching due to larger devices, and speed, which in contrast demands small transistor dimensions.

6

Conclusions

Geheimnisvoll am lichten Tag
Läßt sich Natur des Schleiers nicht berauben,
Und was sie deinem Geist nicht offenbaren mag,
Das zwingst du ihr nicht ab mit Hebeln und mit Schrauben.

FAUST, Der Tragödie erster Teil

The central theme throughout this thesis is the achievement of high accuracy, which is one of the two competing design challenges that have to be faced in A/D converter development. The second challenge is speed. While the conversion rate can be increased by utilizing new technologies, better architectures and more power consumption, the possibilities are more limited when it comes to increasing the accuracy.

In addition, the ongoing miniaturization and the integration of entire systems on a single chip demand the co-integration of analog front-ends together with digital signal processing circuits. Therefore, A/D converters have to be integrated in standard digital CMOS processes

to be compatible with the digital designs. So the traditional, but cost-intensive solutions to increase the accuracy, such as laser trimming, cannot be applied any more because they require special analog processes. The accuracy of A/D converters therefore must be achieved by using the matching capabilities of active or passive elements, such as resistors or capacitors, if no special on-chip trimming or calibration circuits are provided. But even this choice of elements is likely to be limited when using a modern deep submicron process where it is not guaranteed that precise capacitors or resistors are available, as such devices are not needed in digital circuits.

An architecture that provides high matching accuracy when using only MOS transistors is therefore needed. A promising approach is an R-2R-type ladder structure that uses MOSFETs rather than resistors. This circuit is investigated in detail in this thesis.

In a first step, the current division principle, which forms the basis of the operation of the MOSFET-only ladder, is examined. The physical background is discussed and the mechanism of the current division is analyzed. This analysis shows in which operation regions of the MOS transistor the principle is valid and which effects limit the performance. It is of special importance to understand these second-order effects because they decrease the accuracy of the current division. Based on the physical description, these effects are discussed and design guidelines are given to minimize the negative influence on the accuracy. This examination leads to the schematic of the MOSFET-only ladder which, similar to its resistive counterpart, produces a number of binary weighted currents that can be used for D/A converter implementations. Due to the simple and highly regular structure which only needs four transistors per bit, the MOSFET-only ladder yields a very compact layout which is advantageous from a matching point of view.

An important issue when using an R-2R-like structure in applications that require high accuracy is the sensitivity of the circuit to offset between its output terminals. Any offset disturbs the series and parallel

connection of the individual ladder elements and thus causes an error of the output current. This error depends on the digital code applied to the ladder and is therefore signal-dependent, which in turn causes distortion. To calculate the error induced by the output offset, the value of the output resistance of the ladder must be known for each digital code. A general description of the output resistance is developed for resistive R-2R ladders and the applicability of these calculations to MOSFET-only ladders is discussed. Based on the resulting equations the maximum offset voltage for a given tolerable current error is calculated.

Another important aspect of the MOSFET-only ladder is the speed that can be achieved with this structure. The analysis carried out on a one-bit ladder yields a dependence of the settling time on the square of the channel length of the ladder transistors. A trade-off has to be found between good matching accuracy that demands large transistor geometry and a sufficiently high speed, which is only achievable with short channel lengths.

To obtain good matching accuracy, special attention has to be paid to the layout. As mentioned above, due to the structure of the ladder circuit, the layout can be carried out very compact and highly regular. To maximize the matching performance, a one-dimensional common-centroid structure with the most significant bit in the middle of the ladder has been used. Furthermore, dummy structures ensure that the environment is identical for all transistors.

To prove the functionality and the matching performance, pure ladder structures have been implemented in a 1 µm CMOS process. The measurements show that MOS transistors are capable of a matching accuracy of more than 11 bit.

After having verified the good static accuracy of the MOSFET-only ladder, the structure is used as a D/A converter in a successive-approximation A/D converter to prove its dynamic performance. The most important building block of a successive-approximation A/D converter in terms of linearity is the D/A converter. Therefore, this architecture is well suited to test the accuracy of the MOSFET-only ladder.

When using the ladder as a D/A converter, its output currents have to be extracted and made available to the subsequent circuits. This current extraction must be accomplished without compromising the ladder accuracy. As mentioned above, the key requirement is a low offset voltage between the terminals of the ladder. A second, equally important prerequisite is a low input impedance. The regulated cascode circuit is shown to be a good choice for this purpose, especially to provide a very low input impedance without compromising other requirements. The properties of the regulated cascode are examined, the requirements for the operational amplifier that performs the actual regulation are calculated, and design guidelines are given.

The second building block in a successive-approximation A/D converter that influences its dynamic performance is the comparator. In this implementation the input to the comparator consists of two currents rather than voltages due to the output signal of the D/A converter being a current. Because all fast comparator structures only accept voltages as inputs, a non-linear current-to-voltage conversion circuit is provided at the comparator inputs. Due to the non-linearity, small currents are converted to comparatively large input voltages to increase the sensitivity of the comparator, while the input voltage for large currents is reduced to clamp the voltage swing and thus increase speed. The core voltage comparator is a regenerative cross-coupled CMOS latch with a special two-phase clocking scheme to reduce offset.

The A/D converter is implemented in a 1 μm digital CMOS process with two layers of metal and one layer of polysilicon. No special passive devices are available. The resolution of the A/D converter has been chosen as 10 bit, and the conversion speed is 200 ksample/s. An excellent total harmonic distortion (THD) of −79 dB has been measured. The effective number of bits (ENOB) is 9.4 bit at a conversion rate of 100 ksample/s. The silicon area including pads is 5.14 mm^2, and the power consumption from a 5 V supply is 12 mW.

The measurements of the A/D converter show that the MOSFET-only ladder is also capable of high performance in dynamic operation,

not only statically. Therefore, the next step in the development is to increase the conversion speed to several Msample/s. Various options exist to accomplish this goal. One straightforward solution is to improve the successive-approximation algorithm. With this measure, however, the conversion rate can be increased by a marginal factor only. To push up the speed by an order of magnitude or more, other solutions must be found. Sticking with the existing successive-approximation design and using a number of converters in parallel increases the conversion speed effectively, but also the area and power consumption and is only advisable if the required conversion speed is so high that it cannot be achieved otherwise.

For the given accuracy and conversion speed, the pipelined architecture is the best solution. As with the successive-approximation A/D converter, the linearity of the pipelined converter is determined by the D/A converter, which is the most important building block of a pipeline stage. Starting from the flash converter architecture, the development of the pipelined A/D converter and the role of each building block are discussed. Of special importance are the errors stemming from the individual building blocks of a pipeline stage, such as offsets, gain errors and non-linearities of the A/D subconverter. The digital error correction that is implemented in virtually all pipelined converters is discussed. The capability to correct several types of errors is examined, and the non-linearity of the pipelined A/D converter that is caused by those errors that cannot be corrected is calculated. Finally, the choice of the stage resolution is discussed, taking into account speed, accuracy, area and power consumption requirements.

A stage resolution of four bit has been selected for the implementation of the pipelined A/D converter. The architecture of one stage is discussed in detail. Due to using the MOSFET-only ladder and due to the limitations imposed by the $0.25\,\mu m$ CMOS process chosen, the block diagram of one stage deviates to some extent from the standard solution used in many implementations. In particular, the sample-and-hold stage and the interstage amplifier have been separated, which leads

to an elegant solution to convert the ladder output current back to the voltage domain. The overall design decisions as well as the timing of the entire A/D converter are presented.

The D/A converter also uses a regulated cascode stage to extract the ladder current. Due to the increased accuracy, however, the ladder demands a maximum offset voltage of 45 μV. This offset performance is achieved with a special gain-enhanced auto-zero circuit. An in-depth analysis of the requirements of the regulated cascode structure and the operational amplifier is given. The amplifier also uses regulated cascodes for gain enhancement which has the additional benefit that the gain requirements and speed considerations can be separated.

The A/D subconverter in each stage is fully differential, as is the entire pipeline. The analysis of the flash converter structure includes reference ladder bowing, which has to be taken into consideration to optimize the power consumed by the reference ladder string. The fully differential comparators consist of a differential-difference amplifier (DDA) front-end and the same type of core comparator as used in the successive-approximation converter.

The sample-and-hold circuit enables the pipelined operation. A comparatively simple architecture can be chosen due to the fact that no gain has to be provided. The sizing of the hold capacitors is discussed, and the design of the switches, their impact on accuracy and stability is examined. The operational amplifier used in the sample-and-hold circuit is discussed. The load of the amplifier, which is capacitive as well as resistive, requires a second amplifier stage to achieve high gain in spite of the resistive load.

The pipelined A/D converter is implemented in a 0.25 μm technology with one layer of polysilicon and six layers of metal. The supply voltage is 2.5 V. Four stages with four bits per stage yield a total resolution of 13 bit. The maximum sampling rate is 5 Msample/s. A total harmonic distortion of −70 dB has been measured at the maximum conversion speed and an input frequency of 2.09 MHz. The spurious-free dynamic range is 76 dB. The differential non-linearity (DNL)

shows 13 bit performance without trimming. The chip area including pads is 9.63 mm^2, and the power consumption is 117.5 mW.

The measurements presented in this thesis show that the MOSFET-only ladder is capable of high matching accuracy, statically as well as dynamically. The simple structure and small geometry make it a versatile building block that can be employed in any system where a D/A converter is needed, not only in different A/D and D/A converter architectures, but also e. g. for offset trimming of amplifiers or comparators. The measured accuracy of more than 11 bit demonstrates that the matching performance of MOS transistors is as good as what can be achieved with precise passive elements. The pipelined A/D converter design in a 0.25 µm technology shows that this is also true for a modern deep submicron process.

The MOSFET-only nature of the ladder is another advantage compared to passive solutions. This type of ladder can be implemented in any standard digital CMOS process without the need of additional analog options such as a second layer of oxide to obtain precise capacitors or a special high-ohmic polysilicon or metallic layer to implement accurate resistors. As already mentioned, the possibility to implement analog circuits in a standard digital CMOS process is gaining more and more importance, especially for A/D converters. Only if the same technology is used for the analog front-end and the digital circuitry, a co-integration on the same chip is possible.

Both A/D converters presented in this thesis meet the requirements of co-integration to be used in a larger system. But especially the design of the pipelined A/D converter shows the challenges that have to be faced in analog design when demands of the digital circuits cause a migration to smaller and smaller feature size.

Different aspects have to be taken into account when using a deep submicron technology for analog design. The reason for digital designs to move towards smaller feature sizes is the higher speed that can be achieved with these processes. The increased speed is undoubtedly an

advantage for analog design as well. The other side of the coin, however, is the low power supply voltage that accompanies a deep submicron technology.

The most obvious disadvantage of a reduced power supply voltage is the decrease of the dynamic range due to the reduction of the maximum signal amplitude. The 0.25 µm process used for the pipelined A/D converter has a supply voltage of only 2.5 V, compared to 5 V for a 1 µm CMOS process. Therefore, the maximum possible signal amplitude is reduced by a factor of two. Going from a 0.25 µm technology to a 0.18 µm process further decreases the supply voltage to 1.8 V. To maximize the dynamic range fully differential structures have to be used throughout the entire signal path, at the cost of increased complexity and higher power consumption.

A second effect of the low supply voltage is the reduced voltage headroom above and below the signal voltage range. This creates problems in the design of the output stages of operational amplifiers and buffers. Some circuits that have been widely used before, such as source follower output stages, cannot be implemented anymore due to the lack of voltage headroom and have to be replaced by different, more complex solutions.

The small feature size of modern processes not only makes high-speed circuits possible, but also has an important drawback for analog circuits. To achieve high bandwidth, the channel length of the transistors has to be small. This, however, drastically decreases the gain that can be realized in a single amplifier stage. To achieve the required accuracy, on the other hand, a certain minimum gain has to be provided. While, e. g., a Miller-compensated two-stage operational amplifier consisting of seven or eight transistors could be used in a 1 µm technology to achieve a gain of 60 dB to 70 dB, regulated cascode gain-boosted amplifiers have to be employed in a deep submicron process to achieve the same gain. Although the resulting operational amplifier is substantially faster than its counterpart in a 1 µm technology, the complexity is also increased considerably.

The observations made above show a general trend in analog design. Deep submicron technologies allow to increase the circuit speed without problems, but the debt has to be paid manifold at the accuracy side. To achieve the same or even higher accuracy with a deep submicron process as before with 0.5 µm and 1 µm technologies gets more and more difficult. In particular the circuit complexity increases rapidly due to the lower supply voltage and due to the reduced gain, as discussed above. Due to the features of deep submicron processes, the main challenge in analog design more and more is to achieve high accuracy rather than high speed.

At the bottom line, the problems and solutions mentioned above show that the design of A/D converters is far from being taken care of, although the basic principles and many of the architectures still used today have been published forty to fifty years ago. New technologies are being invented that demand new faster A/D converters that are more accurate than their predecessors, and with new, smaller IC technologies even more problems for analog designs will arise. This thesis is meant to be a small step in that direction.

Bibliography

Wer kann was Dummes, wer was Kluges denken,
Das nicht die Vorwelt schon gedacht?

MEPHISTOPHELES, Der Tragödie zweiter Teil

[1] B. M. Gordon, "Linear electronic analog/digital conversion architectures, their origins, parameters, limitations, and applications," *IEEE Trans. Circuits Syst.*, vol. CAS-25, no. 7, pp. 391–418, July 1978.

[2] R. Staffin and R. D. Lohmann, "Signal amplitude quantizer." U.S. Patent 2 869 079, Jan. 1959.

[3] J. J. Fraschilla, R. D. Caveney, and R. M. Harrison, "High-speed analog-to-digital converter and method therefor." U.S. Patent 3 597 761, Aug. 1971.

[4] B. M. Gordon and E. T. Colton, "Signal conversion apparatus." U.S. Patent 2 997 704, Aug. 1961.

[5] D. H. Sheingold, Ed., *Analog-Digital Conversion Handbook*. Englewood Cliffs: Prentice-Hall, 1986.

[6] S. H. Lewis and P. R. Gray, "A pipelined 5-Msample/s 9-bit analog-to-digital converter," *IEEE J. Solid-State Circuits*, vol. SC-22, no. 6, pp. 954–961, Dec. 1987.

[7] M. de Wit, K.-S. Tan, and R. K. Hester, "A low-power 12-b analog-to-digital converter with on-chip precision trimming," *IEEE J. Solid-State Circuits*, vol. 28, no. 4, pp. 455–461, Apr. 1993.

[8] A. N. Karanicolas, H.-S. Lee, and K. L. Bacrania, "A 15-b 1-Msample/s digitally self-calibrated pipeline ADC," *IEEE J. Solid-State Circuits*, vol. 28, no. 12, pp. 1207–1215, Dec. 1993.

[9] M. K. Mayes and S. W. Chin, "Monolithic low-power 16-bit 1MSample/s self-calibrating pipeline ADC," in *ISSCC Dig. Tech. Papers*, San Francisco, CA, Feb. 1996, pp. 312–313.

[10] T.-H. Shu, B.-S. Song, and K. Bacrania, "A 13-b 10-Msample/s ADC digitally calibrated with oversampling delta-sigma converter," *IEEE J. Solid-State Circuits*, vol. 30, no. 4, pp. 443–452, Apr. 1995.

[11] U.-K. Moon and B.-S. Song, "Background digital calibration techniques for pipelined ADC's," *IEEE Trans. Circuits Syst. II*, vol. 44, no. 2, pp. 102–109, Feb. 1997.

[12] S.-U. Kwak, B.-S. Song, and K. Bacrania, "A 15-b, 5-Msample/s low-spurious CMOS ADC," *IEEE J. Solid-State Circuits*, vol. 32, no. 12, pp. 1866–1875, Dec. 1997.

[13] J. Yuan and C. Svensson, "A 10-bit 5-MS/s successive approximation ADC cell used in a 70-MS/s ADC array in 1.2-μm CMOS," *IEEE J. Solid-State Circuits*, vol. 29, no. 8, pp. 866–872, Aug. 1994.

[14] R. Wittmann, W. Schardein, B. J. Hosticka, G. Burbach, and J. Arndt, "Trimless high precision ratioed resistors in D/A and A/D converters," *IEEE J. Solid-State Circuits*, vol. 30, no. 8, pp. 935–939, Aug. 1995.

[15] L. A. Singer and T. L. Brooks, "A 14-bit 10-MHz calibration-free CMOS pipelined A/D converter," in *Symp. VLSI Circuits Dig. Tech. Papers*, Honolulu, Hawaii, June 1996, pp. 94–95.

[16] J. L. McCreary, "Matching properties, and voltage and temperature dependence of MOS capacitors," *IEEE J. Solid-State Circuits*, vol. SC-16, no. 6, pp. 608–616, Dec. 1981.

[17] J.-B. Shyu, G. C. Temes, and F. Krummenacher, "Random error effects in matched MOS capacitors and current sources," *IEEE J. Solid-State Circuits*, vol. SC-19, no. 6, pp. 948–955, Dec. 1984.

[18] K. R. Lakshmikumar, R. A. Hadaway, and M. A. Copeland, "Characterization and modeling of mismatch in MOS transistors for precision analog design," *IEEE J. Solid-State Circuits*, vol. SC-21, no. 6, pp. 1057–1066, Dec. 1986.

[19] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1440, Oct. 1989.

[20] M. J. M. Pelgrom and M. Vertregt, "CMOS technology for mixed signal ICs," *Solid-State Electron.*, vol. 41, no. 7, pp. 967–974, July 1997.

[21] P. E. Allen and D. R. Holberg, *CMOS Analog Circuit Design*. New York: Oxford University Press, 1987.

[22] A. B. Grebene, *Bipolar and MOS Analog Integrated Circuit Design*. New York: John Wiley & Sons, 1984.

[23] T. Miki, Y. Nakamura, M. Nakaya, S. Asai, Y. Akasaka, and Y. Horiba, "An 80-MHz 8-bit CMOS D/A converter," *IEEE J. Solid-State Circuits*, vol. SC-21, no. 6, pp. 983–988, Dec. 1986.

[24] C. A. A. Bastiaansen, D. W. J. Groeneveld, H. J. Schouwenaars, and H. A. H. Termeer, "A 10-b 40-MHz 0.8-μm CMOS current-output D/A converter," *IEEE J. Solid-State Circuits*, vol. 26, no. 7, pp. 917–921, July 1991.

[25] J. Bastos, M. Steyaert, and W. Sansen, "A high yield 12-bit 250-MS/s CMOS D/A converter," in *Proc. IEEE 1996 Custom Integrated Circuits Conf.*, San Diego, CA, May 1996, pp. 431–434.

[26] R. J. Van de Plassche, "Dynamic element matching for high-accuracy monolithic D/A converters," *IEEE J. Solid-State Circuits*, vol. SC-11, no. 6, pp. 795–800, Dec. 1976.

[27] K. Bult and G. J. G. M. Geelen, "An inherently linear and compact MOST-only current division technique," *IEEE J. Solid-State Circuits*, vol. 27, no. 12, pp. 1730–1735, Dec. 1992.

[28] Y. P. Tsividis, *Operation and Modeling of the MOS Transistor*. New York: McGraw-Hill, 1987.

[29] H. Wallinga and K. Bult, "Design and analysis of CMOS analog signal processing circuits by means of a graphical MOST model," *IEEE J. Solid-State Circuits*, vol. 24, no. 3, pp. 672–680, June 1989.

[30] J. E. Schroeder and R. S. Muller, "IGFET analysis through numerical solution of poisson's equation," *IEEE Trans. Electron Devices*, vol. ED-15, no. 12, pp. 954–961, Dec. 1968.

[31] G. Baccarani, M. R. Wordeman, and R. H. Dennard, "Generalized scaling theory and its application to a ¼ micrometer MOSFET design," *IEEE Trans. Electron Devices*, vol. ED-31, no. 4, pp. 452–462, Apr. 1984.

[32] E. Pfrenger, P. Picard, and F. von Sichart, "A companding D/A converter for a dual-channel PCM codec," in *ISSCC Dig. Tech. Papers*, San Francisco, CA, Feb. 1978, pp. 186–187.

[33] G. Merckel, J. Borel, and N. Z. Cupcea, "An accurate large-signal MOS transistor model for use in computer-aided design," *IEEE Trans. Electron Devices*, vol. ED-19, no. 5, pp. 681–690, May 1972.

[34] F. M. Klaassen and W. C. J. de Groot, "Modelling of scaled-down MOS transistors," *Solid-State Electron.*, vol. 23, no. 3, pp. 237–242, Mar. 1980.

[35] T. Poorter and J. H. Satter, "A D.C. model for an MOS-transistor in the saturation region," *Solid-State Electron.*, vol. 23, no. 7, pp. 765–772, July 1980.

[36] C. G. Sodini, P.-K. Ko, and J. L. Moll, "The effect of high fields on MOS device and circuit performance," *IEEE Trans. Electron Devices*, vol. ED-31, no. 10, pp. 1386–1393, Oct. 1984.

[37] C. M. Hammerschmied and Q. Huang, "Design and implementation of an untrimmed MOSFET-only 10-bit A/D converter with –79 dB THD," *IEEE J. Solid-State Circuits*, vol. 33, no. 8, pp. 1148–1157, Aug. 1998.

[38] Q. Huang, C. Menolfi, and C. Hammerschmied, "A MOSFET-only interface for integrated flow sensors," in *Proc. Int. Symposium on Circuits and Systems ISCAS*, Atlanta, GA, May 1996, vol. 4, pp. 372–375.

[39] R. van de Plassche, *Integrated Analog-to-Digital and Digital-to-Analog Converters*. Dordrecht: Kluwer Academic Publishers, 1994.

[40] M. G. Bellanger, G. Bonnerot, and P. Caniquit, "Specification of A/D and D/A converters for FDM telephone signals," *IEEE Trans. Circuits Syst.*, vol. CAS-25, no. 7, pp. 461–467, July 1978.

[41] P. H. Saul, "Successive approximation analog-to-digital conversion at video rates," *IEEE J. Solid-State Circuits*, vol. SC-16, no. 3, pp. 147–151, June 1981.

[42] R. J. van de Plassche and H. J. Schouwenaars, "A monolithic 14 bit A/D converter," *IEEE J. Solid-State Circuits*, vol. SC-17, no. 6, pp. 1112–1117, Dec. 1982.

[43] P. R. Gray and D. A. Hodges, "All-MOS analog-digital conversion techniques," *IEEE Trans. Circuits Syst.*, vol. CAS-25, no. 7, pp. 482–489, July 1978.

[44] K. R. Laker and W. M. C. Sansen, *Design of Analog Integrated Circuits and Systems*. New York: McGraw-Hill, 1994.

[45] E. Säckinger and W. Guggenbühl, "A high-swing, high-impedance MOS cascode circuit," *IEEE J. Solid-State Circuits*, vol. 25, no. 1, pp. 289–298, Feb. 1990.

[46] K. Bult and 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, no. 6, pp. 1379–1384, Dec. 1990.

[47] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*. New York: John Wiley & Sons, third ed., 1993.

[48] D. A. Freitas and K. W. Current, "CMOS current comparator circuit," *Electron. Lett.*, vol. 19, no. 17, pp. 695–697, Aug. 1983.

[49] E. Seevinck, P. J. van Beers, and H. Ontrop, "Current-mode techniques for high-speed VLSI circuits with application to current sense amplifier for CMOS SRAM's," *IEEE J. Solid-State Circuits*, vol. 26, no. 4, pp. 525–536, Apr. 1991.

[50] P. A. Crolla, "A fast latching current comparator for 12-bit A/D applications," *IEEE J. Solid-State Circuits*, vol. SC-17, no. 6, pp. 1088–1094, Dec. 1982.

[51] T. N. Blalock and R. C. Jaeger, "A high-speed clamped bit-line current-mode sense amplifier," *IEEE J. Solid-State Circuits*, vol. 26, no. 4, pp. 542–548, Apr. 1991.

[52] G. M. Yin, F. Op't Eynde, and W. Sansen, "A high-speed CMOS comparator with 8-b resolution," *IEEE J. Solid-State Circuits*, vol. 27, no. 2, pp. 208–211, Feb. 1992.

[53] P. J. Crawley and G. W. Roberts, "High-swing MOS current mirror with arbitrarily high output resistance," *Electron. Lett.*, vol. 28, pp. 361–363, Feb. 1992.

[54] C. Hammerschmied and Q. Huang, "A MOSFET-only, 10b, 200kSample/s A/D converter capable of 12b untrimmed linearity," in *ISSCC Dig. Tech. Papers*, San Francisco, CA, Feb. 1997, pp. 132–133.

[55] J. Doernberg, H.-S. Lee, and D. A. Hodges, "Full-speed testing of A/D converters," *IEEE J. Solid-State Circuits*, vol. SC-19, no. 6, pp. 820–827, Dec. 1984.

[56] W. D. Reeve, *Subscriber Loop Signaling and Transmission Handbook*. Piscataway: IEEE Press, 1995.

[57] W. Goralski, *ADSL and DSL Technologies*. New York: McGraw-Hill, 1998.

[58] K. Bacrania, "A 12-bit successive-approximation-type ADC with digital error correction," *IEEE J. Solid-State Circuits*, vol. SC-21, no. 6, pp. 1016–1025, Dec. 1986.

[59] I. J. Dedic, N. C. Amos, M. J. King, W. G. Schofield, and A. K. Kemp, "A 16b 100ksample/s 2.7V 25mW ADC/DSP/DAC-based analog signal processor in 0.8μm CMOS," in *ISSCC Dig. Tech. Papers*, San Francisco, CA, Feb. 1997, pp. 96–97.

[60] W. G. Schofield, I. J. Dedic, and A. K. Kemp, "A 16 bit 500ks/s 2.7V 5mW ADC/DAC in 0.8μm CMOS using error-correcting successive approximation," in *Proc. ESSCIRC'97*, Southampton, Sept. 1997, pp. 76–79.

[61] C. M. Webster and D. T. Comer, "An enhanced successive-approximation circuit for A/D conversion," in *Proc. IEEE 1993 Custom Integrated Circuits Conf.*, San Diego, CA, May 1993, pp. 28.2.1–28.2.4.

[62] S. M. Bhandari and S. Aggarwal, "A successive double-bit approximation technique for analog/digital conversion," *IEEE Trans. Circuits Syst.*, vol. 37, no. 6, pp. 856–858, June 1990.

[63] C. S. G. Conroy, D. W. Cline, and 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, no. 4, pp. 447–454, Apr. 1993.

[64] K. Nagaraj, H. S. Fetterman, J. Anidjar, S. H. Lewis, and R. G. Renninger, "A 250-mW, 8-b, 52-Msample/s parallel-pipelined A/D converter with reduced number of amplifiers," *IEEE J. Solid-State Circuits*, vol. 32, no. 3, pp. 312–320, Mar. 1997.

[65] W. C. Black, Jr. and D. A. Hodges, "Time interleaved converter arrays," *IEEE J. Solid-State Circuits*, vol. SC-15, no. 6, pp. 1022–1029, Dec. 1980.

[66] A. G. F. Dingwall, "Monolithic expandable 6 bit 20 MHz CMOS/SOS A/D converter," *IEEE J. Solid-State Circuits*, vol. SC-14, no. 6, pp. 926–932, Dec. 1979.

[67] J. G. Peterson, "A monolithic video A/D converter," *IEEE J. Solid-State Circuits*, vol. SC-14, no. 6, pp. 932–937, Dec. 1979.

[68] R. J. van de Plassche and R. E. J. van der Grift, "A high-speed 7 bit A/D converter," *IEEE J. Solid-State Circuits*, vol. SC-14, no. 6, pp. 938–943, Dec. 1979.

[69] A. G. F. Dingwall and V. Zazzu, "An 8-MHz CMOS subranging 8-bit A/D converter," *IEEE J. Solid-State Circuits*, vol. SC-20, no. 6, pp. 1138–1143, Dec. 1985.

[70] S. Sutarja and P. R. Gray, "A pipelined 13-bit 250-ks/s 5-V analog-to-digital converter," *IEEE J. Solid-State Circuits*, vol. 23, no. 6, pp. 1316–1323, Dec. 1988.

[71] B.-S. Song, M. F. Tompsett, and K. R. Lakshmikumar, "A 12-bit 1-Msample/s capacitor error-averaging pipelined A/D converter," *IEEE J. Solid-State Circuits*, vol. 23, no. 6, pp. 1324–1333, Dec. 1988.

[72] S. H. Lewis, H. S. Fetterman, G. F. Gross, R. Ramachandran, and T. R. Viswanathan, "A 10-b 20-Msample/s analog-to-digital converter," *IEEE J. Solid-State Circuits*, vol. 27, no. 3, pp. 351–358, Mar. 1992.

[73] D. A. Mercer, "A 14-b, 2.5 MSPS pipelined ADC with on-chip EPROM," *IEEE J. Solid-State Circuits*, vol. 31, no. 1, pp. 70–76, Jan. 1996.

[74] I. Mehr and L. Singer, "A 55-mW, 10-bit, 40-Msample/s nyquist-rate CMOS ADC," *IEEE J. Solid-State Circuits*, vol. 35, no. 3, pp. 318–325, Mar. 2000.

[75] O. A. Horna, "A 150 Mbps A/D and D/A conversion system," *Comsat Tech. Rev.*, vol. 2, no. 1, pp. 39–72, Spring 1972.

[76] S. H. Lewis, "Optimizing the stage resolution in pipelined, multistage, analog-to-digital converters for video-rate applications," *IEEE Trans. Circuits Syst. II*, vol. 39, no. 8, pp. 516–523, Aug. 1992.

[77] D. W. Cline and P. R. Gray, "A power optimized 13-b 5 Msample/s pipelined analog-to-digital converter in 1.2 μm CMOS," *IEEE J. Solid-State Circuits*, vol. 31, no. 3, pp. 294–303, Mar. 1996.

[78] T. B. Cho and P. R. Gray, "A 10 b, 20 Msample/s, 35 mW pipeline A/D converter," *IEEE J. Solid-State Circuits*, vol. 30, no. 3, pp. 166–172, Mar. 1995.

[79] W.-C. Song, H.-W. Choi, S.-U. Kwak, and B.-S. Song, "A 10-b, 20-Msample/s low-power CMOS ADC," *IEEE J. Solid-State Circuits*, vol. 30, no. 5, pp. 514–521, May 1995.

[80] A. M. Abo and P. R. Gray, "A 1.5-V, 10-bit, 14.3-MS/s CMOS pipeline analog-to-digital converter," *IEEE J. Solid-State Circuits*, vol. 34, no. 5, pp. 599–606, May 1999.

[81] Y.-M. Lin, B. Kim, and 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, no. 4, pp. 628–636, Apr. 1991.

[82] R. Gregorian and G. C. Temes, *Analog MOS Integrated Circuits for Signal Processing*. New York: John Wiley & Sons, 1986.

[83] E. A. Vittoz, "The design of high-performance analog circuits on digital CMOS chips," *IEEE J. Solid-State Circuits*, vol. SC-20, no. 3, pp. 657–665, June 1985.

[84] G. Wegmann, E. A. Vittoz, and F. Rahali, "Charge injection in analog MOS switches," *IEEE J. Solid-State Circuits*, vol. SC-22, no. 6, pp. 1091–1097, Dec. 1987.

[85] M. Degrauwe, E. Vittoz, and I. Verbauwhede, "A micropower CMOS-instrumentation amplifier," *IEEE J. Solid-State Circuits*, vol. SC-20, no. 3, pp. 805–807, June 1985.

[86] T. Burger and Q. Huang, "On the optimum design of regulated cascode operational transconductance amplifiers," in *Proc. 1998 Int. Symposium on Low Power Electronics and Design*, Monterey, CA, Aug. 1998, pp. 203–208.

[87] B. Y. Kamath, R. G. Meyer, and P. R. Gray, "Relationship between frequency response and settling time of operational amplifiers," *IEEE J. Solid-State Circuits*, vol. SC-9, no. 6, pp. 347–352, Dec. 1974.

[88] T. Miki, H. Kuono, T. Kumamoto, Y. Kinoshita, T. Igarashi, and K. Okada, "A 10-b 50 MS/s 500-mW A/D converter using a differential-voltage subconverter," *IEEE J. Solid-State Circuits*, vol. 29, no. 4, pp. 516–522, Apr. 1994.

[89] K. Y. Kim, N. Kusayanagi, and A. A. Abidi, "A 10-b, 100-MS/s CMOS A/D converter," *IEEE J. Solid-State Circuits*, vol. 32, no. 3, pp. 302–311, Mar. 1997.

[90] Ø. Moldsvor and G. S. Østrem, "An 8-bit, 200 MSPS folding and interpolating ADC," *Analog Integrated Circuits and Signal Processing*, vol. 15, no. 1, pp. 37–47, Jan. 1998.

[91] B. Razavi and B. A. Wooley, "A 12-b 5-MSample/s two-step CMOS A/D converter," *IEEE J. Solid-State Circuits*, vol. 27, no. 12, pp. 1667–1678, Dec. 1992.

[92] T. Shih, L. Der, S. H. Lewis, and P. J. Hurst, "A fully differential comparator using a switched-capacitor differencing circuit with common-mode rejection," *IEEE J. Solid-State Circuits*, vol. 32, no. 2, pp. 250–253, Feb. 1997.

[93] E. Säckinger and W. Guggenbühl, "A versatile building block: The CMOS differential difference amplifier," *IEEE J. Solid-State Circuits*, vol. SC-22, no. 2, pp. 287–294, Apr. 1987.

[94] D. G. Haigh and B. Singh, "A switching scheme for switched capacitor filters which reduces the effect of parasitic capacitances associated with switch control terminals," in *Proc. Int. Symposium on Circuits and Systems ISCAS*, Newport Beach, CA, May 1983, vol. 2, pp. 586–589.

[95] R. Hogervorst, J. P. Tero, R. G. H. Eschauzier, and J. H. Huijsing, "A compact power-efficient 3 V CMOS rail-to-rail input/output operational amplifier for VLSI cell libraries," *IEEE J. Solid-State Circuits*, vol. 29, no. 12, pp. 1505–1513, Dec. 1994.

[96] J. H. Huijsing, R. Hogervorst, and K.-J. de Langen, "Low-power low-voltage VLSI operational amplifier cells," *IEEE Trans. Circuits Syst. I*, vol. 42, no. 11, pp. 841–852, Nov. 1995.

List of Symbols

Wie anders wirkt dies Zeichen auf mich ein!

FAUST, Der Tragödie erster Teil

α	Parameter of the Kaiser-Bessel window
α_{jk}	Sign coefficient
β	Transconductance parameter, equals $\mu C'_{\mathrm{ox}}(W/L)$
γ	Body-effect coefficient
δ	Fraction of one LSB
$\mathcal{E}_{\mathrm{c}}$	Critical lateral field in the channel
$\mathcal{E}_x$	Lateral field in the channel
ζ	Proportionality factor
ζ_1	Proportionality factor
ζ_2	Proportionality factor
μ	Surface mobility
μ_{eff}	Effective mobility
σ	Standard deviation
σ^2	Variance

τ	Settling time constant
τ_{reg}	Time constant of comparator regeneration
$\Phi 1$	Clock phase one
$\Phi 2$	Clock phase two
ϕ_F	Fermi potential of extrinsic silicon
ϕ_t	Thermal voltage
ψ_s	Surface potential
ω_2	Non-dominant pole
ω_d	Dominant pole
ω_{GBW}	Gain-bandwidth product of an amplifier
ω_s	Pole caused by a switch
A	Amplifier gain
ΔA	Gain error
A_{aux}	Gain from an auxiliary input to the output
A_{CL}	Ideal closed-loop gain
A_{id}	Ideal amplifier gain
A_{main}	Gain of the main amplifier
A_{min}	Minimum required amplifier gain
A_S	Source area
b_i	Bit i of a digital code
C	Capacitance
C_A	Capacitance at the mid-point node of one ladder slice

C_{BP}	Bottom-plate capacitance
C_{db}	Drain-bulk capacitance
C_{eff}	Effective capacitance
C_{gd}	Gate-drain capacitance
C_{GDO}	Gate-drain overlap capacitance
C_{gs}	Gate-source capacitance
C_H	Hold capacitance
C_i	Lumped capacitance at node i
c_i	Code center of code i
C_{in}	Input capacitance of an amplifier
C_J	Bulk junction capacitance
C_{JSW}	Sidewall bulk junction capacitance
C_L	Load capacitance of a single-stage OTA
C'_{ox}	Gate oxide capacitance per unit area
C_s	Part of node capacitance independent of transistor dimensions
C_{sb}	Source-bulk capacitance
e_i	Code edge of code i
f_{CLK}	Clock frequency
G_m	Transconductance of a single-stage OTA
g_m	Transconductance of an MOS transistor
ΔI	Current error or difference
$I(x)$	Inversion layer current at position x

I_{BIAS}	Bias current
I_D	Drain current
ΔI_D	Drain current change due to current division
I_{diff}	Diffusion current in the inversion layer of an MOS transistor
I_{drift}	Drift current in the inversion layer of an MOS transistor
I_{dump}	Dump current of the ladder
I_{in}	Input current
I_{in+}	Positive input current
I_{in-}	Negative input current
I_j	Current at ladder output for bit j
I_{jk}	Part of current I_j which is caused by voltage source V_k
I_{out}	Output current of the ladder
I_{REF}	Reference current
I_{sw}	Current peak value due to ladder switching
k	Boltzmann constant
L	Transistor length
l_d	Length of source or drain diffusion area
L_{eff}	Effective transistor length
m	Number of stages of a pipelined A/D converter
n	Number of bits
n_s	Number of bits of subsequent stages in a pipelined A/D converter

p	Pole of a transfer function
P_S	Source perimeter
q	Magnitude of electronic charge
Q'_I	Inversion layer charge per unit area
R	Unit resistance of an R-2R ladder or a transresistance stage
r	Small-signal output resistance
r_{in}	Small-signal input resistance
R_{jk}	Part of the output resistance of an R-2R ladder
R_L	Load resistance
R_{min}	Minimum output resistance of an R-2R ladder
R_o	Output resistance of a single-stage OTA
r_{on}	On-resistance of a switch
R_{out}	Output resistance of an R-2R ladder
R_{sp}	Resistance of a non-terminated R-2R ladder
$R_\square$	Sheet resistance
s	Complex variable in the frequency domain
s_n	Code at which the minimum output resistance occurs
T	Absolute temperature
t	Time variable
ΔV	Voltage error or difference
V_A	Mid-point voltage of one ladder slice
V_c	Voltage across a hold capacitor

V_{CM}	Common-mode voltage
v_d	Drift velocity
V_{DD}	Power supply voltage
V_{DS}	Drain-source voltage
V_{DSsat}	Saturation voltage of an MOS transistor
V_{FB}	Flat-band voltage
V_{GATE}	Gate voltage applied to the ladder
V_{GB}	Gate-bulk voltage
V_{GS}	Gate-source voltage
V_{in}	Input voltage
$V_{in_{max}}$	Maximum input voltage
V_{in+}	Positive input voltage of an A/D converter
V_{in-}	Negative input voltage of an A/D converter
ΔV_{inj}	Voltage error caused by charge injection
V_k	Voltage source at ladder output for bit k
ΔV_{max}	Maximum allowed voltage error
V_{nn}	Inverting input of the inverting port of a DDA
V_{np}	Non-inverting input of the inverting port of a DDA
V_{off}	Offset voltage
V_{out}	Output voltage
V_{out+}	Positive output voltage
V_{out-}	Negative output voltage

V_{pn}	Inverting input of the non-inverting port of a DDA
V_{pp}	Non-inverting input of the non-inverting port of a DDA
V_{R+}	Positive reference voltage of a flash A/D converter
V_{R-}	Negative reference voltage of a flash A/D converter
V_{REF}	Reference voltage
V_{SB}	Source-bulk voltage
V_T	Extrapolated threshold voltage of an MOS transistor
V_{T0}	Zero-bias threshold voltage of an MOS transistor
W	Transistor width
W_{opt}	Optimum transistor width
x	Horizontal position in the MOS transistor channel
Z_{out}	Output impedance

Index

Es trägt Verstand und rechter Sinn
Mit wenig Kunst sich selber vor;
Und wenn's euch Ernst ist, was zu sagen,
Ist's nötig, Worten nachzujagen?

FAUST, Der Tragödie erster Teil

Page numbers given in italic type (e. g. '*125*') refer to pages where a figure illustrates an important aspect of the headword in question.

Epigraphs

Acknowledgements

Verzeih, ich kann nicht hohe Worte machen,
Und wenn mich auch der ganze Kreis verhöhnt;
Mein Pathos brächte dich gewiß zum Lachen,
Hättst du dir nicht das Lachen abgewöhnt.

Mephistopheles, Prolog im Himmel, 275–278

Your pardon, if my idiom is lowly,
My eloquence up here would meet with scorn,
Pathos from me would cause you laughter solely,
If laughter weren't a thing you have forsworn.

Mephistopheles, Prologue in Heaven, 275–278

Abstract

Der Teufel hat hier weiter nichts zu sagen.

Mephistopheles, Der Tragödie zweiter Teil, 2. Akt,
Hochgewölbtes enges gotisches Zimmer, 6790

This leaves the devil nothing more to say.

MEPHISTOPHELES, Part Two, Act II,
A High-Vaulted, Narrow Gothic Chamber, 6790

Introduction

Bedenke wohl die erste Zeile,
Daß deine Feder sich nicht übereile!

FAUST, Der Tragödie erster Teil, *Studierzimmer*, 1230–1231

This opening I need to weigh again,
Or sense may suffer from a hasty pen.

FAUST, Part One, *Faust's Study*, 1230–1231

The MOSFET-Only Ladder

So stolz ich bin, muß ich mir selbst gestehn:
Dergleichen hab' ich nie gesehn.

MEPHISTOPHELES, Der Tragödie zweiter Teil, 2. Akt,
Klassische Walpurgisnacht, 7970–7971

Proud as I am, I must admit
I've never seen their like as yet.

MEPHISTOPHELES, Part Two, Act II,
Classical Walpurgis-Night, 7970–7971

A 10 bit Successive-Approximation A/D Converter

Wie ist's? Will's fördern? Will's bald gehn?

FAUST, Der Tragödie erster Teil, *Straße*, 3025

How now? What news? What chances of success?

FAUST, Part One, *A Street*, 3025

Pipelined A/D Converter Architecture

Das heiß' ich endlich vorgeschritten!

MEPHISTOPHELES, Der Tragödie zweiter Teil, 4. Akt,
Hochgebirg, 10067

Now that I reckon's travelling in style!

MEPHISTOPHELES, Part Two, Act IV,
Mountain Heights, 10067

A Pipelined 13 bit A/D Converter

Du wähnst, es füge sich sogleich;
Hier stehen wir vor steilern Stufen.

MEPHISTOPHELES, Der Tragödie zweiter Teil, 1. Akt,
Finstere Galerie, 6193–6194

Out of the blue you think such things appear;
But here with steepest climbing we must reckon.

MEPHISTOPHELES, Part Two, Act I,
A Gloomy Gallery, 6193–6194

Conclusions

Geheimnisvoll am lichten Tag
Läßt sich Natur des Schleiers nicht berauben,
Und was sie deinem Geist nicht offenbaren mag,
Das zwingst du ihr nicht ab mit Hebeln und mit Schrauben.

FAUST, Der Tragödie erster Teil, *Nacht*, 672–675

For Nature keeps her veil inviolate,
Mysterious still in open light of day,
And where the spirit cannot penetrate
Your screws and irons will never make a way.

FAUST, Part One, *Night*, 672–675

Bibliography

Wer kann was Dummes, wer was Kluges denken,
Das nicht die Vorwelt schon gedacht?

MEPHISTOPHELES, Der Tragödie zweiter Teil, 2. Akt,
Hochgewölbtes enges gotisches Zimmer, 6809–6810

For who can think of truth or trash to say,
But someone in the ancient world has thought it?

MEPHISTOPHELES, Part Two, Act II,
A High-Vaulted, Narrow Gothic Chamber, 6809–6810

List of Symbols

Wie anders wirkt dies Zeichen auf mich ein!

FAUST, Der Tragödie erster Teil, *Nacht*, 460

A curious change affects me in this sign.

FAUST, Part One, *Night*, 460

Index

Es trägt Verstand und rechter Sinn
Mit wenig Kunst sich selber vor;
Und wenn's euch Ernst ist, was zu sagen,
Ist's nötig, Worten nachzujagen?

FAUST, Der Tragödie erster Teil, *Nacht*, 550–553

Good sense, Sir, and rightmindedness
Have little need to speak by rule.
And if your mind on urgent truth is set,
Need you go hunting for an epithet?

FAUST, Part One, *Night*, 550–553

The german verses are cited according to:
GOETHE, *Faust*, Kommentiert von Erich Trunz.
München: Ch. Beck, 14. Auflage, 1989.
The translation was done by Philip Wayne, first published in 1949.

Curriculum Vitae

Oct. 1, 1970	Clemens Hammerschmied is born in Salzburg, Austria
Sept. 1980 – June 1988	Akademisches Gymnasium in Salzburg
June 1988	Matura
Oct. 1989 – May 1994	Studies of Electrical Engineering at the Swiss Federal Institute of Technology (ETH) Zurich
May 1994	Diploma degree in Electrical Engineering (Dipl. El.-Ing. ETH)
Since June 1994	Research assistant in the Analog Group of the Integrated Systems Laboratory (IIS) at ETH Zurich. The research topic is MOSFET-only A/D converters.