Digitally Assisted Pipeline ADCs

Theory and Implementation

by
Boris Murmann
and
Bernhard E. Boser
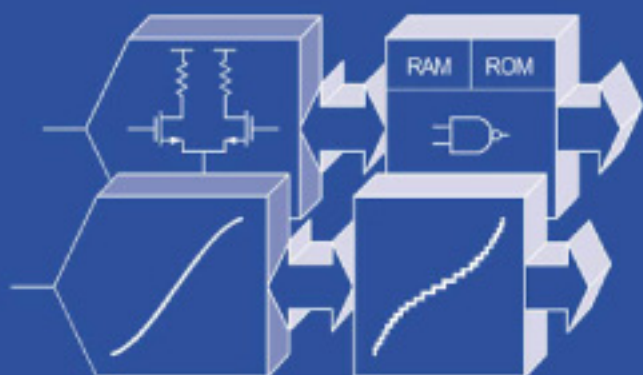
DIGITALLY ASSISTED PIPELINE ADCs

This page intentionally left blank

Digitally Assisted Pipeline ADCs

Theory and Implementation

by

Boris Murmann
Standford University

and

Bernhard E. Boser
University of California, Berkeley

Visit Kluwer Online at: http://kluweronline.com
and Kluwer's eBookstore at: http://ebooks.kluweronline.com

Dedication

To our families.

This page intentionally left blank

Contents

This page intentionally left blank

List of Figures

List of Tables

This page intentionally left blank

Acknowledgments

This page intentionally left blank

Preface

The continued reduction of integrated circuit feature sizes and commensurate improvements in device performance are fueling the progress to higher functionality and new application areas. For example, over the last 15 years, the performance of microprocessors has increased 1000 times. Analog circuit performance has also improved, albeit at a slower pace. For example, over the same period the speed/resolution figure-of-merit of analog-to-digital converters improved by only a factor 10.

Of the many reasons for this disparity between analog and digital circuit performance advances, accuracy requirements stand out as a critical constraint in most analog circuits while being virtually absent in digital designs. Thermal noise, linearity, and matching are distinctly analog circuit problems and require design tradeoffs that invariably lower achievable performance. For example, linearity requirements are usually met with high-gain feedback loops. Unfortunately, this solution also lowers circuit speed and results in elevated noise, reduced signal range, and increased power dissipation.

Technology scaling, while unquestionably advantageous for digital circuits, further exacerbates analog circuit design challenges. While offering increased speed, scaled devices suffer from reduced intrinsic gain, further adding to the design challenge of high-gain feedback loops. Reduced supply voltages lower the ratio of useful signal range to supply, leading to increased power dissipation in noise-limited circuits.

A large range of solutions to overcome these challenges is available to designers, both at the technology and circuits level. At the process level they include high supply options and a choice of transistor threshold voltages. Circuit innovations consist of gain boosting and nested Miller compensation.

While extending the feasibility of analog circuits in scaled technologies with low supply voltages, these techniques come at the cost of a combination of increased process complexity, reduced performance, and added power dissipation.

This book proposes a different approach that takes advantage of the availability of high performance digital processing to relax analog circuit linearity requirements. The use of simple but nonlinear open loop amplification translates into increased analog circuit performance or lower power dissipation. In a careful design that uses a modern process, the area and power penalty of the added digital circuitry is negligible and benefits fully from further technology scaling.

Performance demands and design challenges for analog circuits will continue to increase in the future. This book gives the designer a powerful new tool to meet these demands.

Bernhard E. Boser
Berkeley, January 2004

Chapter 1

INTRODUCTION

1.   MOTIVATION

Enabled by the continuing aggressive scaling of fine line integrated circuit technology, digital signal processing (DSP) and computing have become the main progress drivers in modern electronic systems. With decreasing transistor dimensions, binary computations are performed at lower energy levels and higher speed, resulting in an increasing number of highly sophisticated architectures and algorithms that can be efficiently implemented using digital electronic circuits. In the past decades, this development has led to a continuous doubling of microprocessor performance every 18 months [1].

While purely analog circuits can also benefit from technology scaling, several limitations account for relatively slow performance improvements over time. Most fundamentally, the simultaneous requirement of high speed, low distortion and low noise in the processing of analog signals often translates into poor power efficiency and limited throughput. Furthermore, decreasing supply voltages and reduced intrinsic transistor gain in modern technologies make the design of highly linear, high dynamic range analog building blocks an increasingly challenging task [2].

As a result of these trends, designers lean toward a system partition with a minimum number of virtually unavoidable analog components. Among them is the analog-to-digital converter (ADC) , which is required to interface digital processors to "real life" signals such as radio, image and speech waveforms. Since quantization of continuous amplitude information requires analog operations, ADCs often limit the throughput of DSP based systems. In addition, the fairly high power consumption of today's converters is also

becoming an increasingly severe showstopper. Especially in applications requiring portability, the operating speed of ADCs tends to be set by the allowable power dissipation, rather than the technological limit.

2.   OVERVIEW

This book is concerned with improving the speed and power efficiency of analog-to-digital converters. In particular, we explore the opportunity to overcome analog circuit limitations by incorporating digital domain algorithms into the conversion process. The proposed "digitally assisted" converter makes extensive use of the dense, low cost and low power DSP circuitry available in modern integrated circuit technology.

In recent years, the pipelined ADC in Complementary Metal-Oxide-Silicon (CMOS) technology has become the most popular architecture for high speed Nyquist conversion at medium resolutions of 8-14 bits and conversion speeds ranging from 1-200 Mega-Samples per second (MS/s) . Typical applications include radio receivers and base stations, digital imaging and video, ultra-sound, radar and sonar systems.

In this book, the pipelined ADC topology is used as a vehicle to derive and demonstrate an alternative approach to conventional quantizers that rely on accurate analog signal processing. By delegating many of the precision requirements from the analog to the digital domain, the proposed converter can benefit from technology scaling rather than being impeded by its limitations.

Among the key building blocks in pipelined ADCs are the residue amplifiers that interface successive converter stages. Especially in the converter front-end, these gain elements have to meet very stringent speed, noise and linearity specifications and therefore tend to set the overall power dissipation and attainable speed.

The key feature of this research is a DSP driven technique that alleviates linearity requirements in the analog signal path and thereby helps to break the classical speed-noise-linearity constraint loop. Traditional precision feedback amplifiers are replaced by simple open-loop structures that exhibit superior speed, power efficiency and improved immunity to technology scaling. In the presented proof-of-concept prototype, this approach enables power savings of up to 75% in critical sub-circuits.

Figure 1-1 shows a block diagram of the digitally assisted ADC. A digital post-processor takes the raw, imprecise conversion result and performs the task of identifying and compensating analog domain nonidealities, including mismatch errors and amplifier nonlinearity. In the described converter, the system identification process is based on the

evaluation of the raw code signal statistics, and "blind" in the sense that no precise test signal is superimposed or injected into the analog signal path. The linearization parameters are continuously updated during normal ADC operation to track variations in operating conditions such as temperature and supply voltage.

Digital correction and calibration of analog domain non-idealities is not new. Especially in pipelined ADCs, digital correction [3] and calibration [4] have been used extensively to overcome offset and unit element mismatch errors. However, the characteristic feature of the approach demonstrated here is the extent to which digital compensation is used. Treating distortion in semiconductor circuits as a digital domain problem is the main contribution of this work.

Even though the solution presented is tailored for a specific architecture, most of the general concepts and paradigms can form the basis for similar approaches involving other circuit topologies. Some examples of derivative strategies are summarized in chapter 10.
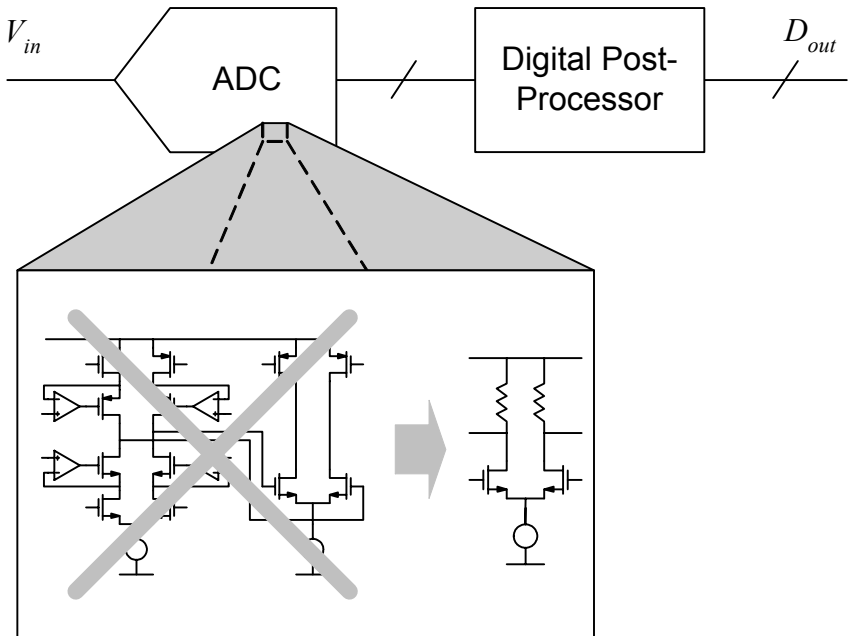
Figure 1-1. System overview.

3.   CHAPTER ORGANIZATION

This book is divided into ten chapters. Chapter 2 reviews ADC figures-of-merit and presents a motivating survey of the trends and impact of technology scaling on ADC performance. It shows that the computing capabilities of digital circuits have outpaced progress in analog-to-digital conversion interfaces by more than two orders of magnitude in the past 15 years.

Chapter 3 revisits the controversial question of the impact of scaling on analog circuit power efficiency, and provides a correction to previous, pessimistic analyses.

Chapter 4 aims to identify opportunities for improving the power efficiency in ADCs. The cost for precise and linear analog signal amplification in terms of power efficiency is evaluated, and serves as the main motivation for the modified, open loop pipelined ADCs discussed in chapter 5.

Chapters 6 and 7 describe the proposed digital post-processing mechanism that compensates for linear and nonlinear pipeline stage non-idealities. The two main elements of the developed scheme are a redundancy-based digital correction mechanism and a statistics based background calibration technique.

Chapter 8 details the implementation of a 12-bit 75 MS/s pipelined ADC [5] that was used to evaluate the proposed concepts. Detailed measurement results confirming the feasibility of the digitally assisted ADC concept are illustrated in chapter 9. Highlights of these results include the digital reduction of the converter's integral nonlinearity error from 18 to less than 0.7 least significant bits (LSBs) .

Chapter 10 contains a summary of this book and presents a proposal for future research and development.

Chapter 2

PERFORMANCE TRENDS

1. INTRODUCTION

In the past decades, "Moore's Law" [6] has governed the revolution in microelectronics. Through continuous advancements in device and fabrication technology, the industry has maintained exponential progress rates in transistor miniaturization and integration density. As a result, microchips have become cheaper, faster, more complex and power efficient.

This chapter surveys the impact of technology scaling on the performance of digital circuits and analog-to-digital interfaces; the focus is placed on the past 15 years, during which CMOS technology has been the most popular technology for a large number of applications.

As shown in the following sections, digital performance metrics have grown faster than relevant metrics in ADCs. The resulting large and growing performance gap is the motivation of this research towards a more "digitally assisted" conversion interface.

In the context of the presented data, it should be noted that an objective comparison of absolute performance metrics over time is difficult. Benchmarks in electronic systems are usually expressed using "figures of merit" that lump several performance characteristics into one number. Finding and assigning an appropriate weight to each of the contributing aspects is challenging, subjective and context dependent. For instance, the trend towards portable, battery-operated equipment has led to a shift in paradigms toward power efficient systems, resulting in a change of constraints and goals over time. This comparative survey aims to illustrate only orders of magnitude in relative performance improvement over time and avoids such second order considerations.

2. DIGITAL PERFORMANCE TRENDS

Digital circuit applications can be regarded as the main driver for semiconductor device scaling. Historically, the development of new CMOS technology generations has been primarily motivated by the rapidly growing demand for high performance in digital microprocessors. Smaller feature sizes result in faster transistor switching speeds and lower energy consumption per binary transition.

While it is clear that technology scaling must eventually come to an end, the current roadmap of the Semiconductor Industry Association (SIA) foresees a continuation of the above trend up until the year 2016, when the physical transistor gate length is expected to reach 9nm [7]. Table 2.1 summarizes the progress in feature size and integration density over the past 15 years [1].

Table 2.1. Moore's Law: Integration density in lead microprocessors.

	1987	2002	Rate of Change
Transistor Gate Length (L)	1 μm	0.13 μm	0.5x every 5 years
Transistors/Die	$\cong$1 Million	$\cong$100 Million	2x every 2.3 years

2.1 Microprocessor Speed

The attainable speed in digital circuits is approximately proportional to the technology feature size. A widely accepted figure of merit for digital circuit speed is the so-called "fan-out of four" (*FO*4) delay [8]. As illustrated in Table 2.2, this metric has been continuously reduced by a factor of two every 5 years, which coincides with the rate of feature size reduction in technology.

Table 2.2. Speed in lead microprocessors.

	1987	2002	Rate of Change
Delay (*FO*4) $\cong 360ps \cdot L/\mu m$ [9]	360ps	47ps	0.5x every 5 years
Clock Speed	20MHz	1.7GHz	2x every 2.3 years
SPECInt 2000 Performance	$\cong$1	$\cong$1000	2x every 1.5 years
MIPS Performance	$\cong$10	$\cong$10,000	2x every 1.5 years

Aside from this raw speed improvement, designers have managed to achieve further performance enhancements both by refining logic gate topologies and by increasing the level of pipelining. Pipelining reduces the number of gate delays between registers and thus improves system throughput. As a result of these factors, clock speed in lead microprocessors has doubled approximately every 2.3 years. This growth is more than twice that of *FO*4 delay.

An additional advantage in microprocessors that adds to the overall computing power is the extensive amount of parallelism feasible in fine line technologies. On top of the quickly growing clock speed, architectural parallelism has led to a net doubling of computing power every 1.5 years. Quantifying the computing power of a microprocessor objectively is difficult and controversial [10]. However, both the hardware-oriented "MIPS" metric and the more accepted computing measure "SPECInt" show this tremendous growth rate (see Table 2.2) [11].

2.2 Microprocessor Power Efficiency

Feature size scaling has decreased the energy per logic transition by 65% in each technology generation [12]. Equivalently, this corresponds to an energy reduction by a factor of two every 1.7 years. This dramatic rate of improvement stems from both smaller capacitance and lower supply voltage, which has quadratic impact on energy.

For high performance microprocessors, however, this advantage is offset by the extra effort spent on pipelining and architectural parallelism to boost computing power. As a result, the power efficiency of lead microprocessors, measured in mW/MIPS has decreased only by about 40% per technology generation (see *Table 2.3*).

Table 2.3. Digital energy/power efficiency.

	1987	2002	Rate of Change
Relative Energy per Transition ($\propto C_{ox}V_{DD}^2$)	1	$1.8 \cdot 10^{-3}$	0.5x every 1.7 years
Lead Microprocessor Power Efficiency	200mW/MIPS	10mW/MIPS	0.5x every 3.4 years

3. ADC PERFORMANCE TRENDS

Analog circuits, including ADCs, have also benefited from the technology scaling that is mostly driven by digital applications. Today's

mainstream CMOS technology has proven to be most suitable for cost-efficient implementation of high-performance data converters, filters and radio frequency transceivers. Recent performance highlights that make ultimate use of the available integration density and speed in CMOS include an 8-bit, 20-GSample/s ADC [13], and 5-GHz transceiver chips for wireless local area networks [14-16].

In the following survey, we will examine the rate of performance growth in ADCs. To capture and compare performance of ADCs, we use a set of commonly used figures of merit. The following section briefly discusses these quantities with respect to their origin and limitations.

3.1   ADC Figure of Merit Considerations

The product of conversion bandwidth and number of effective quantization levels represents the most basic performance metric for ADCs [17]. We define this quantity as

$$FOM1 \quad f_s \cdot 2^{ENOB}, \tag{2-1}$$

where f_s is the sampling rate of the converter and *ENOB* is the effective number of bits given by

$$ENOB \quad \frac{SNDR - 1.76 dB}{6.02 dB}. \tag{2-2}$$

Since the signal-to-noise and distortion ratio (*SNDR*) of a converter usually depends on the frequency of the input signal, this figure of merit must include some fixed condition for the frequency at which *ENOB* was measured. Alternatively, it is common to replace the sampling rate f_s in (2-1) by twice the signal bandwidth for which the peak *ENOB* has dropped by 3dB. This frequency is often referred to as the effective resolution bandwidth (*ERBW*) [17, 18] .

A fundamental issue in the figure of merit described by (2-1) lies in the relative weighting of throughput and accuracy. For instance, the expression implies that a 6-bit converter running at 1GS/s is equally "hard to build" as a 7-bit converter that operates at 500MS/s. While there is no fundamental argument that holds up this exact tradeoff, it is well supported in practice. The survey [17] shows that for every octave increase in bandwidth, the attainable resolution of state-of-the-art ADCs tends to drop by approximately one bit.

A second, commonly used figure of merit that includes the power dissipation of the ADC is the "energy per conversion" figure of merit given by [19]

$$FOM2 \quad \frac{P}{f_s \cdot 2^{ENOB}}.$$ (2-3)

Note that contrary to the standard convention used in figures of merit, a smaller value of this metric indicates better performance.

In *FOM2*, the tradeoff between precision and power is controversial. Equation (2-3) suggests that the power consumption of an ADC should double for each added bit. However, assuming that the ADC is limited by *kT/C* thermal noise, adding an extra bit requires quadrupling the effective capacitance in the converter. This in turn, requires a 4x increase in current and power dissipation to maintain the same speed. Based on this argument, some authors use a figure of merit in which the denominator carries the precision as 2^{2ENOB}. In practice, this modification is overly pessimistic, since almost never all power dissipating circuits are limited by thermal noise. For improved accuracy, one could introduce a fitting parameter in the denominator, such that

$$FOM2^* \quad \frac{P}{f_s \cdot 2^{c \cdot ENOB}},$$ (2-4)

where *c* is a constant that quantifies the tradeoff between power and precision for a specific ADC architecture. Figures of merit of this form have recently been proposed [20]. In practice, however, it turns out that *c*=1 is a sufficiently good choice to compare ADCs over many technology generations, topologies, speeds and resolutions [17]. As a result, (2-3) has evolved as one of the most widely accepted figure of merits for ADCs.

One way to avoid the problem of uncertainty in the exact power-resolution tradeoff is to compare only converters with approximately the same effective resolution. The corresponding quantity is given by

$$FOM3 \quad \frac{P}{f_s}\bigg|_{ENOB \cong fixed}.$$ (2-5)

This figure of merit is most useful when comparing specific implementations of virtually identical converter topologies, e.g. 10-bit

pipelined ADCs. We will use (2-5) in a detailed architecture-specific ADC survey in chapter 3.

In the following sections we use (2-1) and (2-3) for a more general trend survey on the impact of technology scaling on ADCs of all variants.

3.2   ADC Throughput

Figure 2-1 illustrates the trend in ADC throughput since 1987. The performance data for this survey origins from [17][1], augmented with additional data from the International Solid-State Conference (ISSCC) from the years 1999-2003. Each data point in Figure 2-1 corresponds to a specific, single ADC reported in the respective year. An exponential fit to all data points from 1987-2003 shows that the ADC *FOM*1 (equation (2-1)) has doubled only every 6.5 years. A fit to only the peak performance data points in each year yields a slightly faster progress rate of doubling every 4.7 years.

This difference in slopes may be due to the fact that many ADCs are not optimized for peak throughput alone, but also for good power efficiency or other application-specific constraints. Nevertheless, the slow improvement of the peak performance indicates that the progress in conversion interfaces has been lagging that of purely digital circuits discussed in section 2.
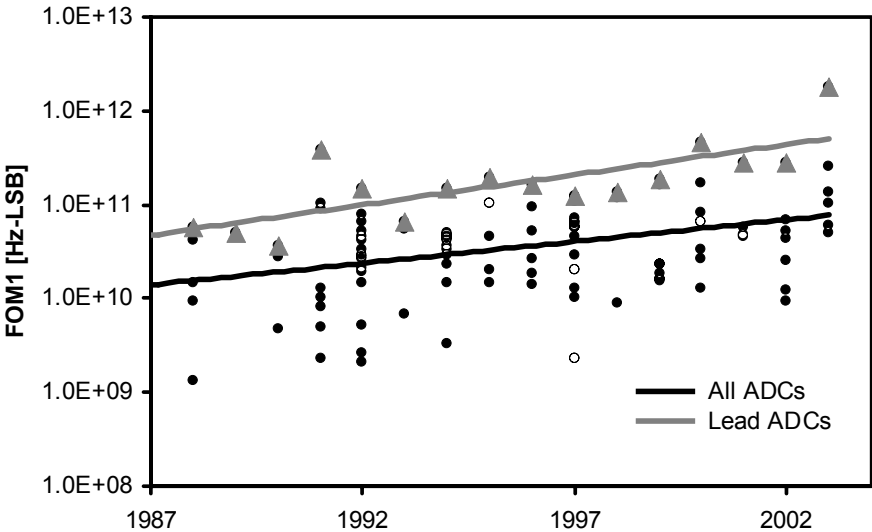
Figure 2-1. ADC performance trend.

---

[1] ADCs using cooled, superconducting devices have been excluded here.

3.3   ADC Energy Efficiency

Using the same source data as in section 3.2, Figure 2-2 shows the development of the energy per conversion figure of merit (equation (2-3)) over time. Again, we perform two distinct fits to the scatter plot. Taking all ADCs into account, *FOM*2 has halved every 2.7 years since 1987, leading to a current state-of-the art value of roughly 3pJ per conversion.

A fit to only the lowest energy parts in each year shows slightly slower progress (0.5x every 3.4 years).   This difference in progress rates between low energy and mainstream ADCs may be due to a general emphasis on low power systems in the 1990s.

3.4   Trend Comparison

It is now interesting to compare the advancements in ADCs to those of digital circuits on a relative scale. Figure 2-3 illustrates the divergence in attainable speed between the two domains.

As explained in section 2.1, microprocessors benefited from the raw improvement in technology speed, and also from aggressively increasing parallelism. The resulting steep progress rate of performance doubling every 1.5 years has created a performance gap of 150x between digital computing power and ADC speed.
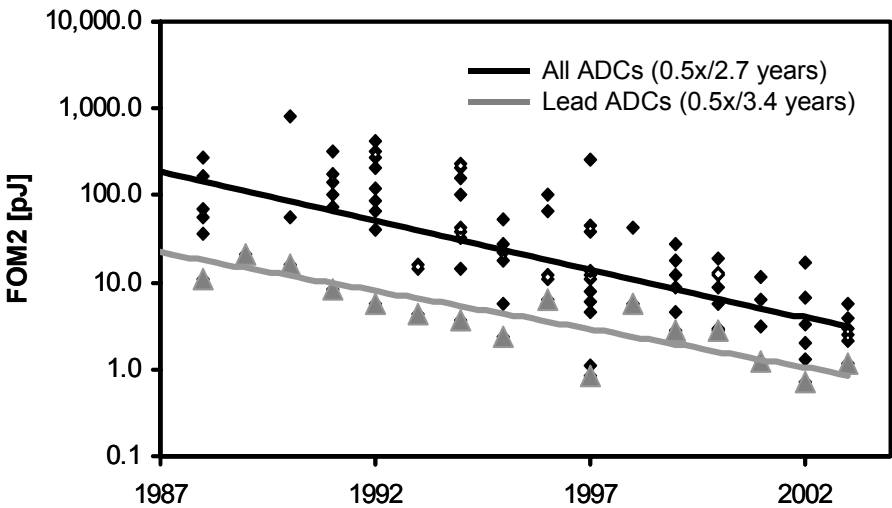
Figure 2-2. ADC energy efficiency trend.
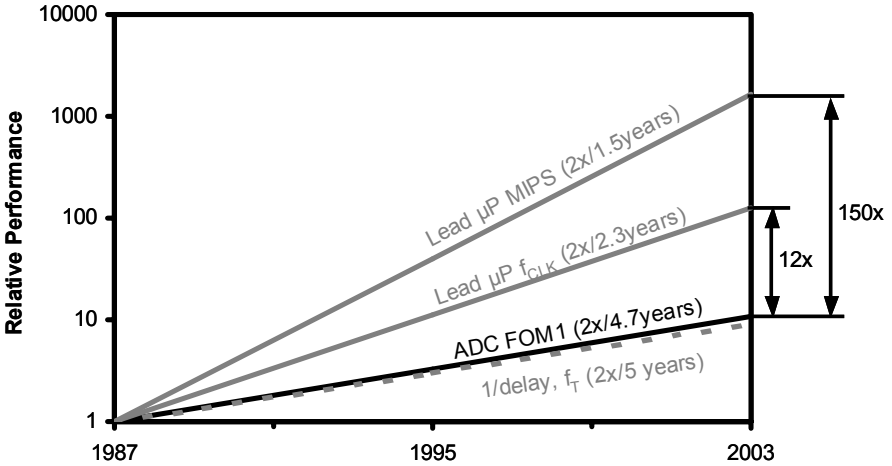
Figure 2-3. Comparison of speed trends: ADCs versus digital.

The situation for energy efficiency is similar. As shown in Figure 2-4, the energy efficiency of logic gates has outperformed the energy per conversion in ADCs by a factor of 14.
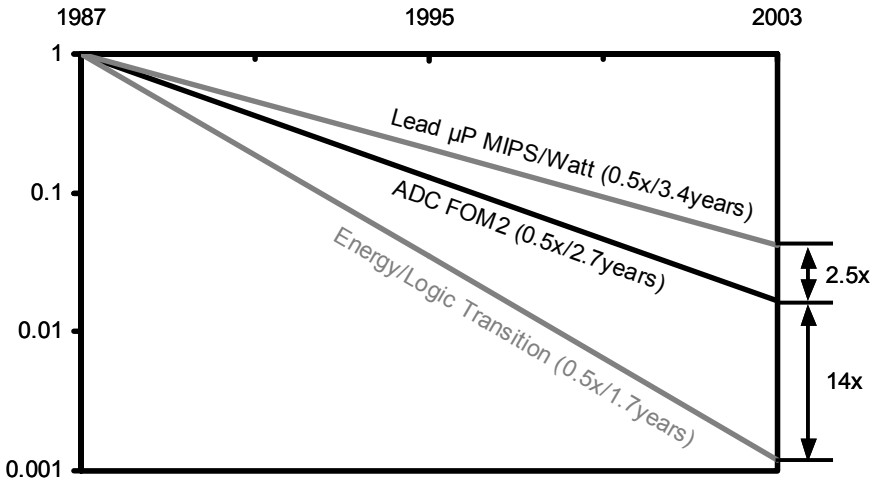
Figure 2-4. Comparison of energy efficiency trends: ADCs versus digital.

It is interesting to note, however, that the overall energy efficiency of lead microprocessors has not improved as fast as that of ADCs. For performance-optimized lead microprocessors, the intrinsic progress in logic gate efficiency is offset by the overhead from architectural parallelism.

Despite this fact, it is clear that there exists a large and growing gap between analog and digital capabilities. Leaving the architectural growth component aside, progress in logic circuits has outpaced ADCs by about 12x in speed (f_{CLK} in Figure 2-3) and 14x in energy efficiency.

To an increasing extent, data converters are the bottleneck of many systems both for throughput and power dissipation. As an example, Figure 2-5 shows a typical mixed-signal application in which both the ADC and digital signal processing backend, consisting of roughly one million logic gates, have been integrated on the same chip. Interestingly, as typical in such applications, the ADC portion (upper right corner) occupies only a small fraction of the die area but consumes more than 50% of the total system power.

Power inefficiency has become one of the most severe showstoppers in the application of ADCs. In many cases, the throughput of ADCs is set by the allowable power dissipation. Figure 2-6 shows several ADC application regimes in the speed/resolution space with contours of equal power consumption.
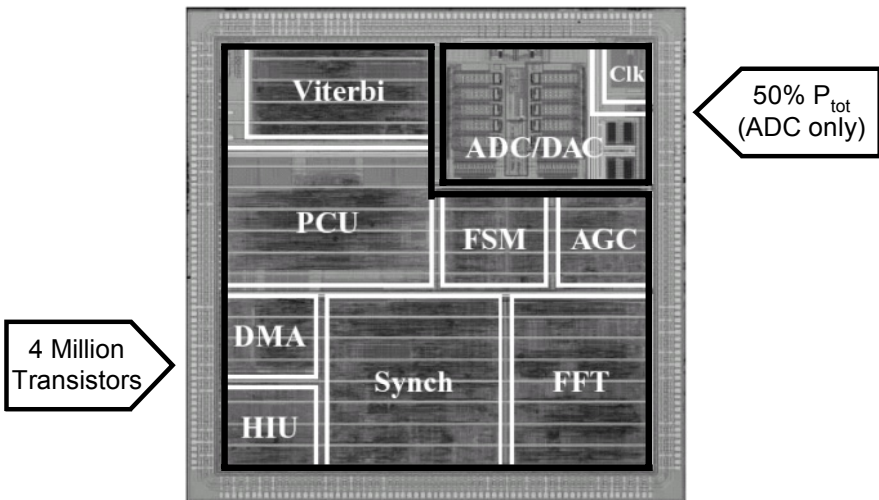
Figure 2-5. Modern ADC application: 802.11 base band processor for wireless networks [21].

With the increasing trend towards battery-powered devices, the power budget of an ADC is usually limited to a fraction of a Watt. As we see from Figure 2-6, this dictates a very strict upper limit in performance that is independent of technology limits.

The large and growing gap between ADC performance and power efficiency, compared with the capabilities of low-power digital devices poses the main motivating question behind this research: How can we use digital circuits to boost the figure of merit in conversion interfaces? The potential advantage of increased "digital assistance" in converters has been recognized and documented in numerous recent publications on the subject (e.g. [22-28]). However, most of the proposed schemes have not yet delivered a significant advantage over "purely analog," optimized ADCs.
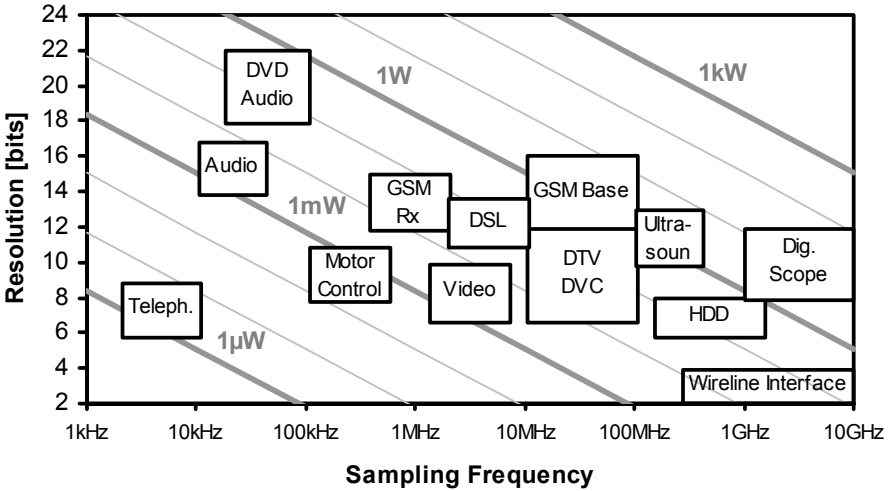
Figure 2-6. ADC applications in the speed/resolution space. The equi-power contours assume *FOM2*=3pJ/conversion.

Chapter 3

SCALING ANALYSIS

1.   INTRODUCTION

For many analog building blocks, including ADCs, it is not clear how power efficiency changes as a function of implementation feature size. Some previously published analyses suggest that there is a detrimental price for implementing high dynamic range functions in a low voltage, deep sub-micron technology [29, 30]. Based on these analyses, the energy figure of merit is bound to deteriorate in fine-line, low-voltage technologies. However, as we have seen in the previous chapter, the migration to finer line widths has not yet caused a reduction in the energy efficiency of ADCs.

The following analysis revisits the controversy over the impact of scaling on analog circuits. The study combines first- and second-order circuit effects and survey data to yield a more refined view that helps explain the trends seen in the previous chapter. The investigation contains three parts:

- A brief summary of CMOS device scaling. How and why are technology parameters varied as channel length decreases?
- Identification and scaling analysis of transistor performance metrics that are important for analog circuits.
- An investigation of how scaling of transistor metrics affects the power efficiency of analog circuits. Here, we distinguish between "matching-limited" and "noise-limited circuits," and focus on representative building blocks of flash- and pipelined ADCs respectively.

2.   BASIC DEVICE SCALING FROM A DIGITAL PERSPECTIVE

From a digital circuit perspective, MOS transistors have been scaled continuously to achieve: (1) higher integration density and reduced cost, (2) higher speed, and (3) lower power consumption. These goals are met by following certain scaling guidelines, which, to first order, have two independent variables: the minimum device feature size, and the supply voltage (V_{DD}).

As explained in [31], the so-called "full scaling approach" attempts to keep electrical fields in the device constant by scaling both voltages and physical dimensions equally. This scaling approach effectively achieves the three scaling goals mentioned above. In practice, however, constant field scaling is not feasible since built-in potentials and the sub-threshold slope (set by kT/q) do not scale with transistor dimensions. Therefore, some form of "general scaling" is usually needed. In this approach, voltages and geometries are reduced by slightly different scaling factors. For each technology generation, the scaling parameters are chosen with the primary objective of maximizing the performance improvement over the previous generation.

One consequence of the general scaling approach, however, is that robustness and reliability tend to trade-off with attainable performance. Some of the resulting issues are:

–   Active power density is steadily rising due to slower V_{DD} scaling relative to dimension scaling.
–   Transistor threshold voltages (V_{TH}) must be scaled down with V_{DD} to prevent performance loss [31]. However, leakage currents increase roughly 10x for every 100mV drop in V_{TH}. This translates into the inability to effectively turn off the device. A minimum allowable V_{TH} of about 0.2V is expected [32].
–   Increased sensitivity to interconnect parasitics. The RC delay of wires has been scaling much slower than device delays [31]. Better interconnect material (e.g. Copper) and improved circuit-level routing solutions have become necessary.

Despite the challenges above, digital circuits are expected to benefit from scaling CMOS technology for at least another five years. Conservative estimates predict that the energy per logic transition will continue to drop until the channel length reaches about 40nm [32].

3. TECHNOLOGY METRICS FOR ANALOG CIRCUITS

Performance metrics for a given technology can be divided into analog and digital parameters. While a digital circuit designer might care mostly about a technology's ring oscillator frequency and energy per logic transition, these parameters have no direct meaning in the context of analog circuits.

In the following sections, we summarize important technology performance parameters from the viewpoint of an analog circuit designer and examine their change with technology scaling. We use qualitative arguments and simulation data from BSIM3v3 models [33] to quantify scaling behavior. Most of the underlying device models were obtained from the MOSIS foundry service web site [34]. For brevity, we restrict the study to four representative technology nodes at 0.5μm, 0.35μm, 0.25μm and 0.18μm. These generations span roughly 7.5 years on the scaling roadmap and are sufficient to predict and analyze general trends.

3.1 Supply Voltage

Signal headroom plays an important role in the design of analog circuits. As supply voltages decrease as dictated by the general scaling approach, many analog functions become harder to implement. For instance, with reduced headroom, it may no longer be feasible to stack transistors in cascode configuration to achieve high output impedance and gain (see e.g. [30]). Another detrimental factor is the achievable dynamic range of the circuit. As the available signal swing scales down by U, noise power in the circuit must be reduced by U^2 to maintain a given dynamic range. This effect is important in noise-limited analog circuits, which are analyzed in more detail in section 5. For further comparison and figure of merit calculations, we use supply voltages from the current and previous technology scaling roadmaps [7] (see Figure 3-1). Over the four technology nodes of interest, supply voltages have been reduced from 5V (0.5μm) to 1.8V (0.18μm).
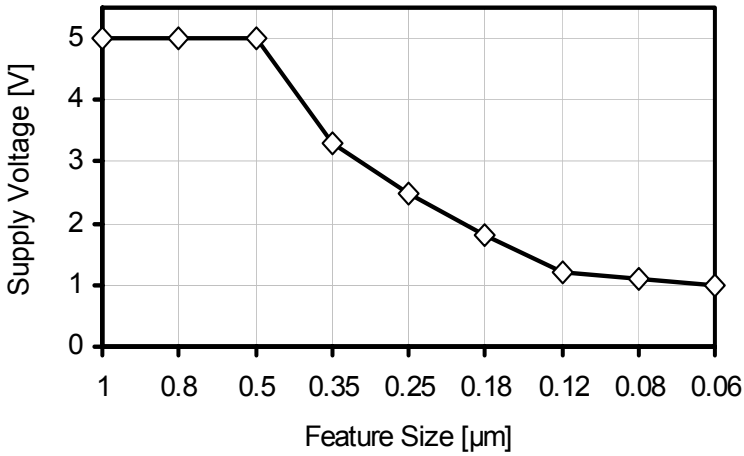
Figure 3-1. Supply voltage scaling.

3.2   Transit Frequency

The transit frequency (f_T) can be regarded as a small-signal, high frequency figure of merit for transistors. At the operating frequency $f=f_T$, a transistor is defined to have unity current gain in a common source configuration with shorted drain. Therefore,

$$f_T \quad \frac{1}{2\pi} \frac{g_m}{C_{gs} + C_{gd}}, \tag{3-1}$$

where g_m is the device's transconductance and C_{gs} and C_{gd} are its gate-source and gate-drain capacitances, respectively. Assuming square law models (see e.g. [35]), f_T is related to device parameters by

$$f_T \cong \frac{1}{2\pi} \frac{\mu \cdot V_{OV}}{L^2}, \tag{3-2}$$

where μ is the channel mobility and V_{OV} is the gate overdrive V_{GS}-V_{TH} of the transistor. Due to short-channel effects such as mobility degradation and velocity saturation, f_T tends to scale by a factor of less than $1/L^2$. Figure 3-2 shows simulation data of NMOS transit frequency for minimum length devices in different technologies versus gate overdrive voltage V_{OV}.
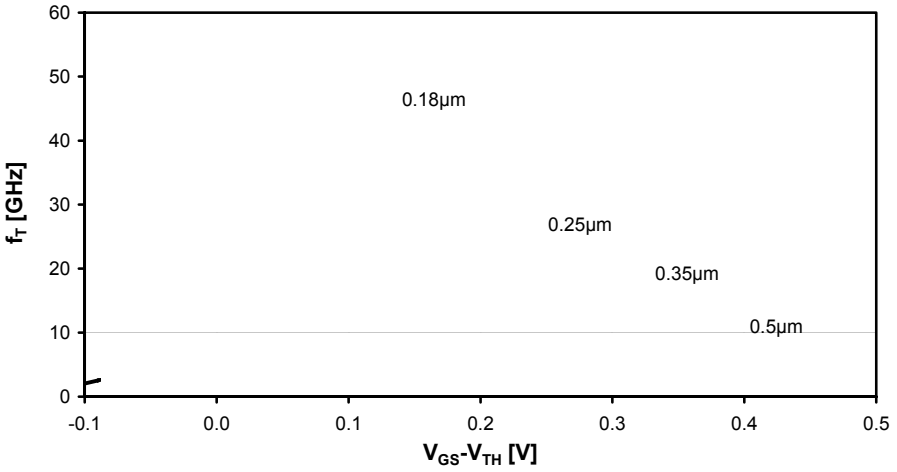
Figure 3-2. NMOS transit frequency.

As we argue later, the available device f_T can be directly related to analog building block bandwidth and is therefore an important metric in deriving figures of merit for analog circuit purposes. As opposed to the drastic drop in supply voltage, availability of transit frequencies approaching 100GHz is a welcome feature for cutting edge analog designs and enables pushing the operating speed.

3.3   Transconductor Efficiency

The transconductor efficiency g_m/I_D quantifies the available device transconductance per current invested. For a square law transistor model, g_m/I_D is given by

$$\eta \quad \frac{g_m}{I_D} \quad \frac{2}{V_{OV}} .\tag{3-3}$$

For practical devices, η is always below the ideal value predicted by (3-3). For very small gate overdrive V_{OV} (<50mV), the device enters a region close to bipolar operation and g_m/I_D is bounded by the value $1/(n{\cdot}kT/q)$, where n is the transistor's sub-threshold slope factor [36]. For large gate overdrive, velocity saturation and mobility degradation cause g_m/I_D to be about 10-20% below the square law estimate. Figure 3-3 shows the transconductor efficiency for the technologies considered here.
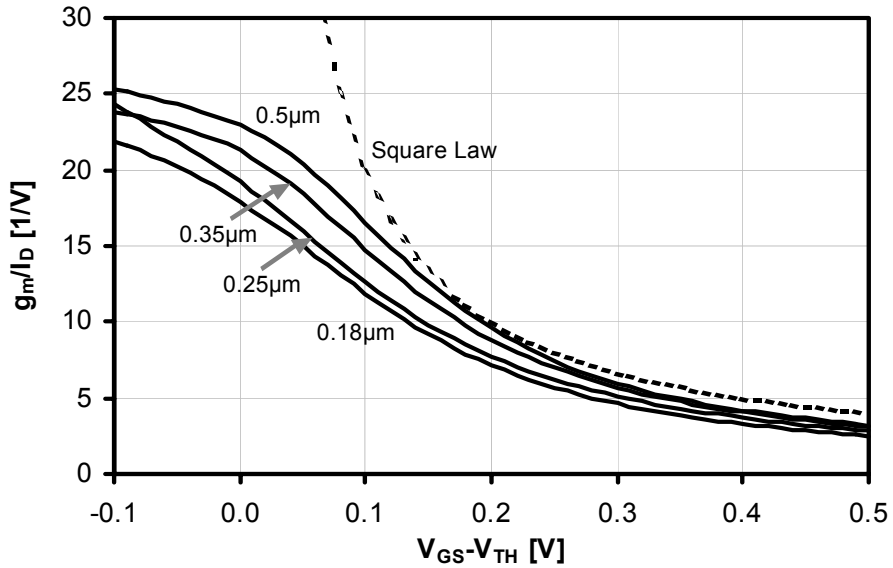
Figure 3-3. Transconductor efficiency versus gate overdrive. The dotted line shows the case for perfect square law devices.

The 0.18μm technology shows the lowest g_m/I_D in all operating regions. This is due to the fact that this technology exhibits the largest sub-threshold slope factor and suffers most from short-channel effects. In future technology generations, enhancements such as strained silicon [37] may help reduce the relative impact of this penalty.

From the perspective of analog circuit design, it is interesting to plot the product of transconductor efficiency and transit frequency. To some degree, this quantity captures the fundamental tradeoff between speed and power and helps to identify reasonable operating regimes for analog transistors. Figure 3-4 shows a plot for the technologies under consideration.

For most technologies, the optimal biasing, i.e. the maximum of $g_m/I_D \cdot f_T$ occurs close to a gate overdrive voltage of 150-200mV. It is interesting to note that the peak is at lower gate overdrive for smaller gate lengths. This trend is explained by the effect of mobility reduction due to increasing vertical electric fields in smaller feature sizes [36].
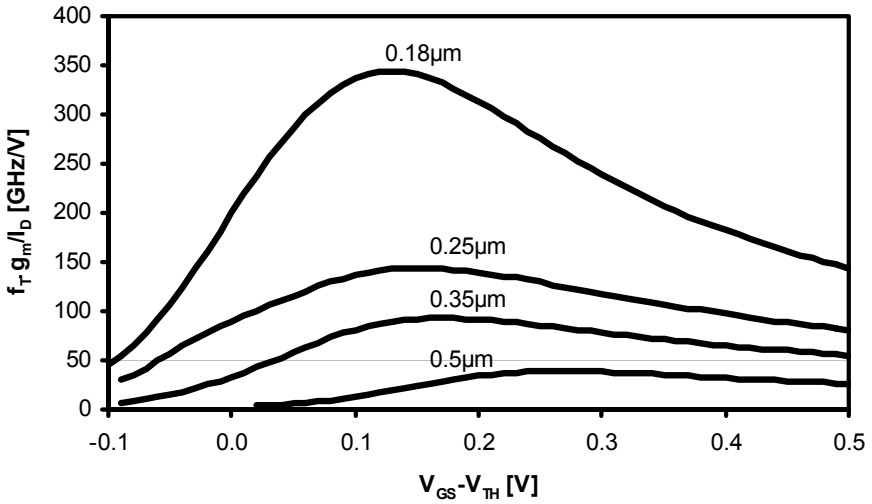
Figure 3-4. Product $g_m/I_D \cdot f_T$.

3.4 Intrinsic Gain

In solid-state transistors, the relationship between control node voltage and device current is highly nonlinear. Linear gain elements are therefore typically implemented using electronic feedback. With feedback, nonlinearities are attenuated by the circuit's loop gain T. In most amplifier configurations, T is given by a product of individual intrinsic transistor gains $(g_m \cdot r_o)$. With decreasing transistor geometries, the intrinsic device gain decreases, which makes it harder to meet minimum loop gain requirements in precision building blocks.

Device physics shows that the decrease in intrinsic gain is due to increased channel length modulation and Drain Induced Barrier Lowering (DIBL) for shorter channels [36]. Figure 3-5 shows intrinsic gain for the different technologies and drain bias for V_{OV}=200mV. Especially critical in the 0.18µm case is the extremely gentle transition to acceptable gain levels. A drain bias of roughly $3 \cdot V_{OV}$=0.6V is required to achieve a device gain of 20. This voltage is a large fraction of the total swing that can be accommodated at V_{DD}=1.8V. Figure 3-6 shows a zoom into the realistic biasing range that can be allocated in today's designs. Just like decreasing V_{DD}, the low intrinsic device gain in short channel technologies can be regarded as a dynamic range penalty.
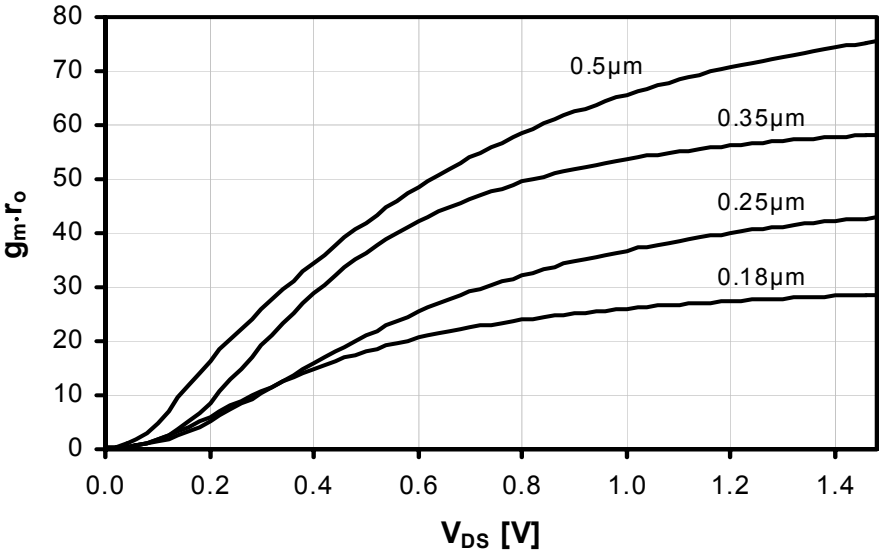
Figure 3-5. NMOS intrinsic device gain at V_{OV}=200mV (minimum channel length).
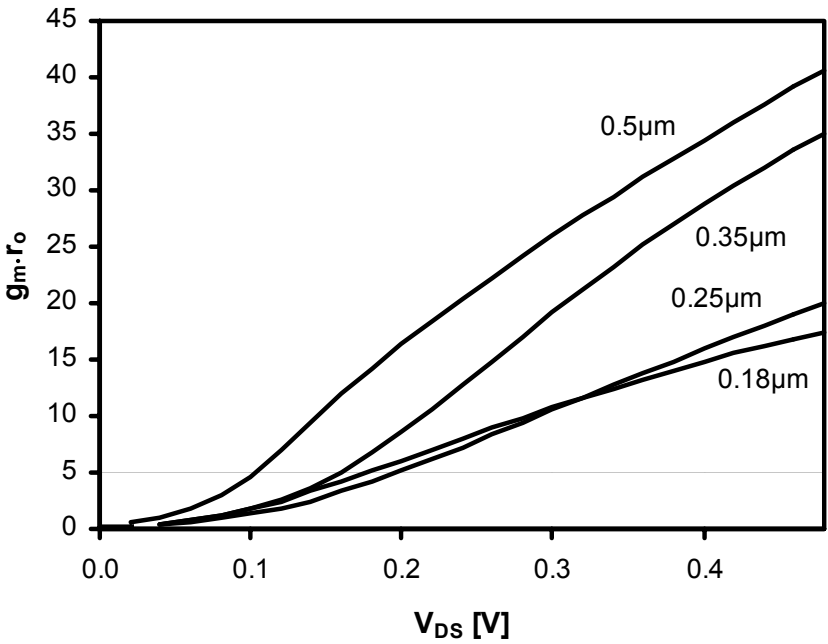
Figure 3-6. NMOS intrinsic device gain at V_{OV}=200mV (Zoom into typical operating region).

3.5   Transistor Matching

Since many analog circuits are based on multiples of supposedly identical devices, matching is often critical.   For certain topologies, matching becomes the bottleneck for attainable accuracy. The mismatch of transistor parameters is also affected by technology scaling. This section provides an introduction to basic matching properties, and their scaling trends.

The most widely accepted description of the variation in some parameter P between two "identical" rectangular devices was first introduced in [38]

$$\sigma^2\left(\Delta P\right) \quad \frac{A_P^2}{W \cdot L} + S_P^2 \cdot D^2 , \tag{3-4}$$

where A_p is an area proportionality constant for parameter P, and S_p describes the variation in P due to spacing. Once the process-dependent constants A_p and S_p have been measured or calculated, this relation can be used to predict matching characteristics of various devices.

Analog circuit designers are normally concerned about transistor current mismatch and/or voltage offset.   For a differential transistor pair with identical size and bias, these quantities are given by

$$\frac{\Delta I_D}{I_D} \quad \frac{\Delta \beta}{\beta} - \frac{g_m}{I_D} \cdot \Delta V_{TH} \tag{3-5}$$

$$\text{and} \quad \Delta V_{GS} \quad \Delta V_{TH} - \frac{I_D}{g_m} \cdot \frac{\Delta \beta}{\beta} , \tag{3-6}$$

where

$$\beta \quad \mu_{eff} \cdot C_{ox} \cdot W / L . \tag{3-7}$$

Due to its random nature, the mismatch is usually described in terms of variance. Using (3-4), the random variations in the threshold voltage and current factor become

$$\mathrm{var}\left(\Delta V_{TH}\right) \cong \frac{A_{VTH}^2}{W \cdot L} \tag{3-8}$$

$$\mathrm{and} \quad \mathrm{var}\left(\frac{\Delta\beta}{\beta}\right) \cong \frac{A_\beta^2}{W \cdot L}. \tag{3-9}$$

These expressions neglect distance effects. In practice, this is a very good assumption for device separation below 200μm [39].

Scaling trends of mismatch can be analyzed by relating fluctuations in device manufacturing to physical device parameters. Threshold voltages are determined mainly by oxide thickness and depletion charge in the channel. Variations in the threshold voltage are caused mostly by the random nature of the ion implantation and diffusion processes, which leave an amount of fixed charges in the depletion region. Assuming that V_{TH} mismatches are due mainly to these random doping fluctuations, one can show that A_{VTH} is directly proportional to oxide thickness [40, 41]. As a result, threshold voltage matching improves with technology scaling.

This is confirmed by Figure 3-7, which shows data for six generations of CMOS technology [42]. Unfortunately, as explained in [42], A_β tends to remain constant with technology scaling. Although ΔV_{TH} has so far been the dominant factor to overall mismatch performance, $\Delta\beta$ is becoming increasingly important. In fine-line technologies, the two mismatch components can be comparable, and both need to be taken into account.
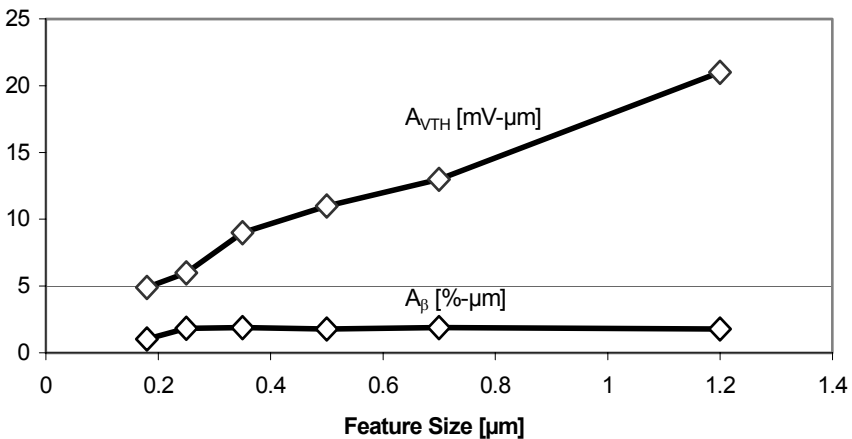
Figure 3-7. Technology scaling trends of A_{VTH} and A_β.

3.6   **Transistor Noise**

In MOS transistors, two significant mechanisms contribute to drain current fluctuations. Flicker noise or 1/f noise is present due to trapping and de-trapping effects at the silicon-oxide interface [43]. Since flicker noise is inversely proportional to transistor gate area, this noise component typically increases with technology scaling. Analog building blocks exhibit different levels of sensitivity to flicker noise depending on their function and application. In the wideband circuits discussed in Sections 4 and 5, flicker noise is usually of minor concern.

The second, more fundamental noise source is thermal noise, whose power spectral density is

$$\overline{i_d^2} \quad \gamma \cdot 4kT \cdot g_m \cdot \Delta f \ . \tag{3-10}$$

For long channel devices, γ=2/3. Recent measurement results show that γ is approximately 1 in 0.18µm technology [44]. This additional noise adds another component to the dynamic range penalty of scaled technologies.

4.   **SCALING IMPACT ON MATCHING-LIMITED CIRCUITS**

Having identified basic device performance scaling trends, we now relate this data to performance of analog building blocks. The following discussion focuses on basic building blocks that comprise ADCs and distinguishes between "matching-limited" circuits discussed in this section and "noise-limited" circuits considered in section 5.

As an example of a matching-limited circuit we study the impact of scaling on flash ADCs. Due to the low resolutions (~4-8 bits), thermal noise tends to be of minor concern in this architecture. However, to achieve high sampling rates, low complexity circuits and small device areas are imperative. For this reason, device matching typically limits the achievable resolution. Like most data converters, flash ADCs exhibit technology-dependent tradeoffs between speed, accuracy, and power consumption. While technology scaling results in the usual short-channel degradations and reduced supply headroom, matching tends to improve. Hence, it is unclear whether smaller feature sizes produce better or worse performance and power efficiency.

4.1   Impact of Mismatch

The achievable resolution of Flash ADCs depends on how accurately the analog input can be compared to a set of incremental reference levels. The general topology for a flash ADC is shown in Figure 3-8.

To alleviate offset requirements, a pre-amplifier usually precedes each comparator. As a result, the offset voltage in each signal path tends to be dominated by the differential pair of the pre-amplifier alone. One way to express the offset voltage of a differential pair is

$$V_{OS} \quad \Delta V_{GS} \quad \Delta V_{TH} - \frac{I_D}{g_m} \cdot \frac{\Delta\beta}{\beta}. \tag{3-11}$$

Equivalently, we may re-write (3-11) in terms of variances

$$\text{var}(V_{OS}) \quad \text{var}(\Delta V_{TH}) + \left(\frac{I_D}{g_m}\right)^2 \text{var}\left(\frac{\Delta\beta}{\beta}\right). \tag{3-12}$$
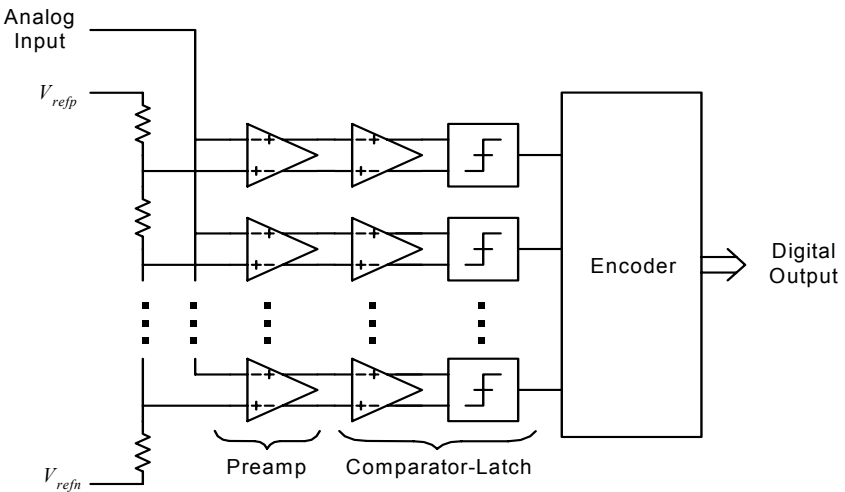
Figure 3-8. Flash ADC block diagram.

Substituting (3-8) and (3-9) into (3-12), we obtain

$$\sigma^2(V_{OS}) \quad \frac{1}{A_{eff}}\left[A_{V\,TH}^2 + A_\beta^2 \cdot \left(\frac{I_D}{g_m}\right)^2 \right], \qquad (3\text{-}13)$$

where A_{eff} is the effective device area, $W \cdot L$.

4.2 Speed Limitations

The conversion speed of a flash ADC is limited mainly by the effective bandwidth of the preamp/comparator. Consider the simple model for pre-amplifier/comparator interface shown in Figure 3-9.

Typically, the preamp input stage provides a voltage gain of approximately 3 (or 2-4). The number of time constants needed to reach some settling accuracy is related to the desired resolution. For instance, in a 6-bit ADC we require $\ln(2^6) \cong 4$ time constants to settle to the desired accuracy. For a given unity gain frequency f_u in the pre-amplifier, the conversion rate is therefore limited to

$$f_s \cong \frac{f_u}{3 \cdot 4} \quad \frac{f_u}{12}. \qquad (3\text{-}14)$$
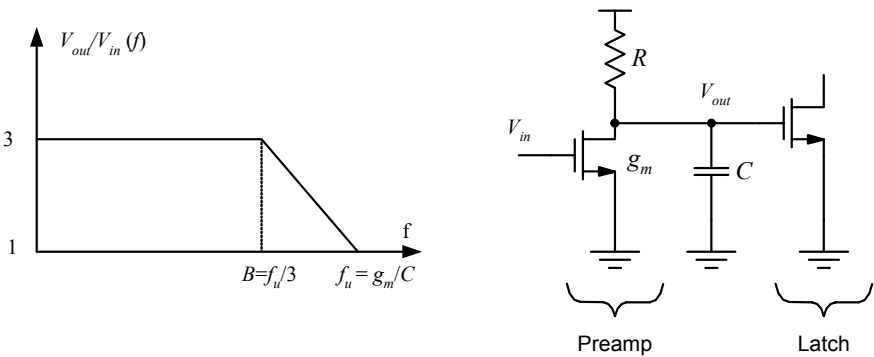
Figure 3-9. Preamp/latch model.

4.3   **Power-Speed-Accuracy Figure of Merit**

A meaningful metric for converters that are not limited by thermal noise is the power per speed-accuracy figure of merit given by [42]

$$FOM_{PSA} \propto \frac{Power}{Speed \cdot Accuracy^2}. \tag{3-15}$$

This quantity has units of energy and indicates how much power must be invested at a given conversion rate to achieve a certain (fixed) resolution.

Power consumption is given by

$$Power \propto I_D \cdot V_{DD}. \tag{3-16}$$

The expression above implies that the circuits are purely class-A, i.e. continuously biased by constant currents. The amount of digital circuitry in flash topologies varies significantly from one implementation to the next. For simplicity, we neglect digital power consumption in this analysis.

The desired resolution translates into a required accuracy. As we argued above, the attainable accuracy here is limited mainly by mismatch. However, another component that affects the achievable resolution is the reference voltage, which is directly related to the supply voltage. Hence, the accuracy term is

$$Accuracy \propto \frac{r \cdot V_{DD}}{\sigma(V_{OS})}, \tag{3-17}$$

where r is the fraction of supply voltage used as the full-scale input range of the converter. For simplicity in this analysis, we assume $r=1$.

The achievable speed of the converter is given by the bandwidth of the pre-amp input stage driving a comparator latch (see Figure 3-9). The unity-gain bandwidth f_u is given by

$$f_u \quad \frac{1}{2\pi} \frac{g_m}{W \cdot \left(c \cdot L_{min} \cdot C'_{gs} + C'_{db}\right)}, \tag{3-18}$$

where C'_{db} and C'_{gs} are the drain-to-bulk and gate-to-source junction capacitance per device width. The constant c relates the device sizes of the two stages and is specific to the topology used. For simplicity in the

following discussion, we assume $c=1$. In practice, this is almost never true. However, since c is roughly independent of technology, it can be safely ignored in a relative scaling analysis.

Combining equations (3-15)-(3-18), we obtain

$$FOM_{PSA} \propto \frac{I_D}{g_m} \cdot \frac{\left(L_{min} \cdot C_{gs}{}' + C_{db}{}' \right)}{V_{DD} \cdot L_{min}} \cdot \left[A_{VT}{}^2 + A_\beta{}^2 \left(\frac{I_D}{g_m} \right)^2 \right]. \qquad (3-19)$$

To isolate different mechanisms of technology scaling, we now first assume constant A_{VT} and A_β for each technology. Figure 3-10 shows the resulting FOM_{PSA} as a function of desired conversion speed f_s. For this graph, g_m/I_D and capacitance values are generated using SPICE simulations. A given f_s determines the required device f_T and also the maximum g_m/I_D (see Figures 3-2 and 3-3).

Under the assumption that matching does not improve, Figure 3-10 shows that each technology becomes better than its predecessor only after a certain frequency threshold, beyond which the older generation has insufficient transistor speed. This trend is explained by the fact that both low V_{DD} and short channels penalize the power through increased accuracy requirements (see (3-19)).
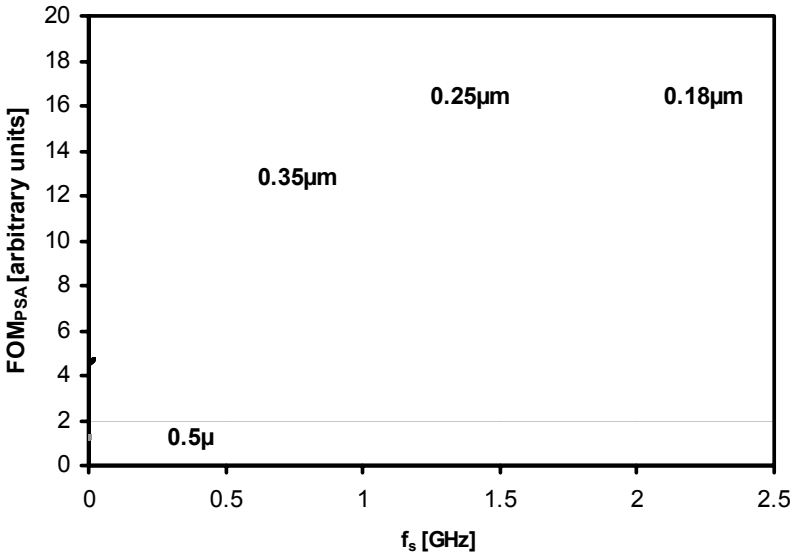
Figure 3-10. Flash ADC energy as a function of sampling rate (assuming constant mismatch factors A_{VTH}, and A_β).

In Figure 3-11, we let the matching coefficients A_{VTH} and A_{β} scale as described in Section 3. As a result, we now see significant merit in scaling, even for moderate speeds. In contrast to Figure 3-10, Figure 3-11 shows that technologies with smaller feature sizes can achieve simultaneous speed and power efficiency improvements.

In order to relate this data to progress over time, we now construct a speed/scaling trajectory. For the four marked data points in Figure 3-11, we assume a typical average flash ADC speed of 350MHz in 0.5µm technology, and a throughput doubling every two process generations (see chapter 2). This choice is somewhat arbitrary, but fairly reasonable. The resulting power efficiency versus feature size is plotted in Figure 3-12.
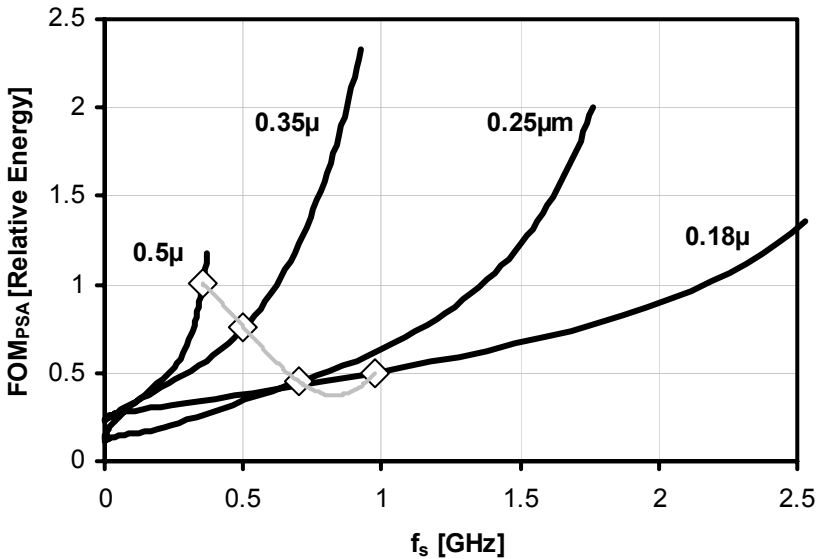
Figure 3-11. Flash ADC energy as a function of sampling rate (assuming improving mismatch factors A_{VT}, and A_{β} with technology).
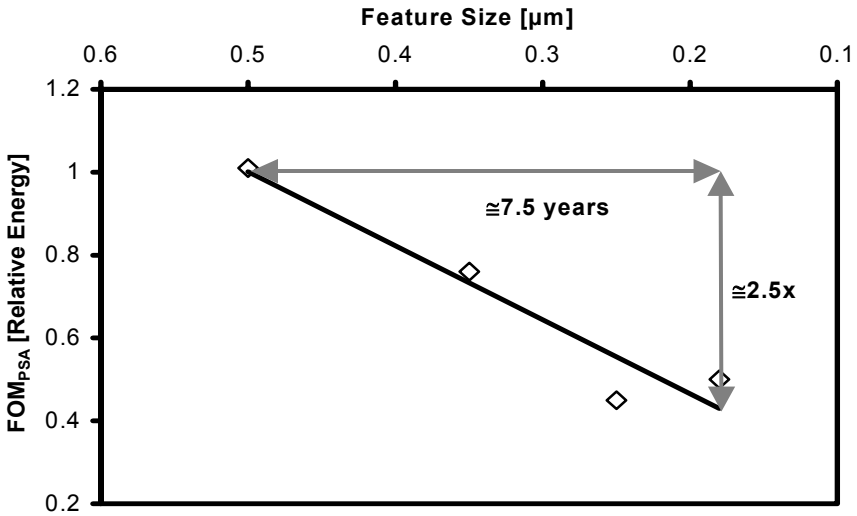
Figure 3-12. Estimated flash ADC energy versus feature size
(from speed trajectory in Figure 3-11).

4.4   **Flash ADC Performance Trends**

It is interesting to compare the result above to published performance data. The data summarized in Table 3-1 is plotted in Figure 3-13 against feature size. Here, we compare only flash ADCs with a fixed resolution of 6 bits, and hence use *FOM*3 as defined in equation (2-5). The linear fits to the data points of Figure 3-12 and Figure 3-13 show a remarkably close energy efficiency improvement rate of roughly 2.5x over the 7.5 years spanned by the four technology nodes under investigation. This corresponds to a 2x energy reduction every 5.7 years.

Note that this progress rate is significantly worse than that seen in the global energy per conversion survey of chapter 2 (*FOM*2 improves 2x every 2.7 years). This observation indicates that energy efficiency of flash ADCs may not scale as well as that of other ADC topologies.

Table 3-1. 6-bit Flash ADC Performance.

Reference	Feature Size [μm]	Year	Speed [MS/s]	Power [mW]	Supply [V]	*FOM3*=Power/Speed [mW/MS/s]
[45]	0.70	1996	175	160	3.3	0.91
[46]	0.60	1999	500	330	3.0	0.66
[47]	0.50	1996	200	110	3.0	0.55
[48]	0.5	1998	350	225	5	0.64
[49]	0.5	1998	400	200	5	0.5
[50]	0.5	1998	200	150	5	0.75
[51]	0.35	1998	400	190	3.0	0.48
[52]	0.35	1999	500	225	3.3	0.45
[53]	0.35	2001	1100	300	3.3	0.27
[54]	0.35	2001	1000	500	3.3	0.5
[55]	0.25	2003	1300	600	1.8	0.46
[56]	0.25	2002	400	150	2.2	0.38
[57]	0.25	2000	700	187	3.3	0.27
[58]	0.25	2000	800	400	3.3	0.50
[59]	0.25	2001	900	450	2.5	0.50
[60]	0.18	2003	2000	310	1.8	0.16
[61]	0.18	2003	400	106	1.8	0.26
[62]	0.18	2002	1600	340	1.9	0.21
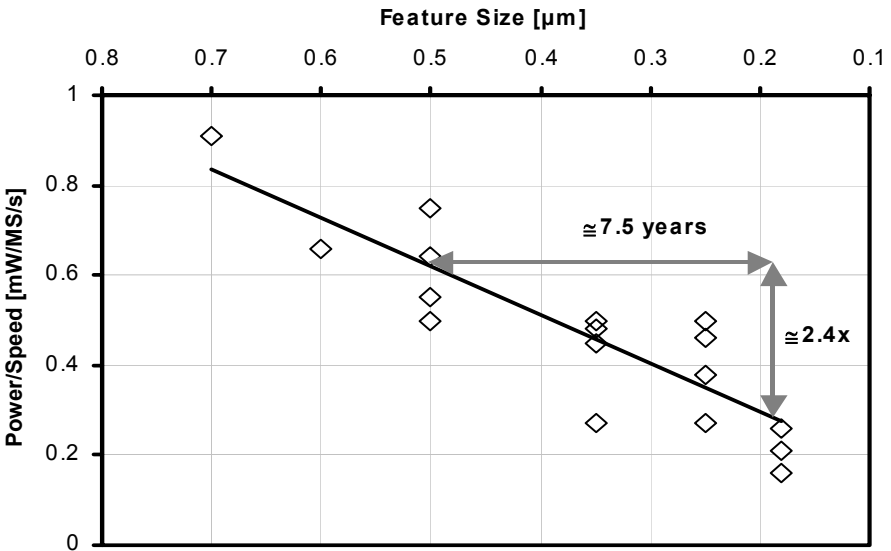
Figure 3-13. Published flash ADC performance vs. technology.

4.5   Discussion

The energy efficiency of flash ADCs in scaled technologies depends strongly on the scaling behavior of matching performance. If matching did not scale, moving to smaller feature sizes would be justified only by a need for a higher speed that is not feasible in a previous technology. Since matching generally improves with technology, we have seen not only higher throughput in flash ADCs, but also improved power efficiency.

Despite the good agreement in the above data, we must be aware of several limitations in the accuracy of this prediction: First, we neglected the digital portion of the ADC. In some designs, the digital circuitry of a flash converter consumes 40-60% of the total power [45, 47]. With the data for energy efficiency of digital circuits from chapter 2 (2x improvement every 1.7 years), this suggests that we should actually see a faster net rate of progress than that seen in Figure 3-12.

Secondly, both our analysis and survey do not take any second order dynamic performance limitations into account. Achieving significantly higher speed in new technologies places stringent requirements on timing and circuit topology, which may adversely affect the complexity and power consumption of the design.

Lastly, the analysis does not conclude topological advancements, such as the use of offset cancellation techniques or interpolation. Increasing design expertise is an important factor in progress, but it is virtually impossible to capture.

Nevertheless, the results above provide good qualitative insight into the scaling behavior of matching-limited circuits and help explain the trends of the past decade.

5.   SCALING IMPACT ON NOISE-LIMITED CIRCUITS

In high-resolution ADCs, the power consumption tends to be set by noise constraints rather than matching. In cases where matching is critical, the desired accuracy is usually achieved through some form of calibration. As an example for a noise-limited circuit, we examine a basic transconductor in feedback configuration. To first order, this circuit resembles the precision amplifiers used in the front-end of sigma-delta and pipelined ADCs.

5.1   **First Order Analysis**

A very basic analysis for noise-limited transconductors was presented in [30]. For a noise limited circuit, an appropriate figure of merit is given by

$$FOM_{PSD} \quad \frac{Power}{Speed \cdot DynamicRange}. \tag{3-20}$$

Consider now the circuit of Figure 3-14 to identify the individual variables of (3-20).

In this circuit, we assume a single transistor amplifier in a "constant" feedback network, i.e. we assume that device loading does not alter the feedback factor F. Furthermore, we assume that the available output swing is proportional to the technology's V_{DD} and the total integrated noise in the circuit is set by the load capacitor C. If we also assume that the transconductor efficiency g_m/I_D is kept constant with technology scaling, we obtain

$$FOM_{PSD} \propto \frac{V_{DD} \cdot I_D}{\dfrac{g_m}{C} \cdot \dfrac{V_{DD}^{\;2}}{kT/C}} \propto \frac{1}{V_{DD}} \tag{3-21}$$

This result states that noise-limited analog power consumption will scale inversely with the technology's supply voltage V_{DD}. For instance, scaling from 0.5μm with V_{DD}=5V to 0.25μm with V_{DD}=2.5V would double power consumption.
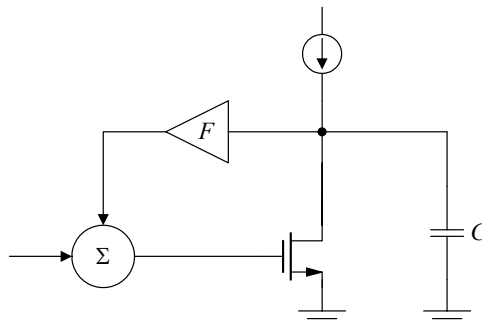
Figure 3-14. Basic amplifier model.

From the trends seen in the previous chapter, it is clear this result overestimates the scaling penalty. In the following sections we will investigate several second order factors that help improve the accuracy of our prediction.

5.2 Modified Analysis

The simple circuit model of Figure 3-14 fails to capture a number of effects that may be significant when trying to predict scaling behavior. In the following sections, we list and examine additional considerations.

5.2.1 Feedback Factor F

As technology scales, capacitive loading of the (capacitive) feedback network by the device C_{gs} decreases. Since circuit speed is proportional to F, improvements in F can translate into lower power at a given speed, and thus help to counteract power increase with scaling. However, in high dynamic range circuits it is usually true that $C_{gs} \ll C_{feedback}$. We therefore consider this effect as insignificant within the scope of this analysis.

5.2.2 Fractional Swing

The peak output swing of the transconductor is more precisely given by

$$Swing \cong V_{DD} - c \cdot V_{OV} \,, \qquad (3\text{-}22)$$

where c accounts for the number of devices that are connected between the supply rails at the output node and an additional V_{DS} margin beyond the minimum value of V_{OV}. Since scaled technologies offer higher transit frequencies, it may be possible to reduce V_{OV} to counteract some of the loss in signal swing due to V_{DD} scaling. However, since this effect is strongly dependent on circuit topology, we will also not consider it further in this analysis.

5.2.3 Transconductor Efficiency g_m/I_D

By the same argument as in 5.2.2, lowering the gate overdrive in scaled technologies may improve g_m/I_D, and translate into power savings at a given operating speed (see Figure 3-3). To examine this effect we plot FOM_{PSD}

using device simulation data in Figure 3-15 with the following considerations:

– The sampling frequency of the converter is linked to the required device f_T through settling and stability constraints. Assuming that the non-dominant amplifier pole occurs around f_T, the attainable loop bandwidth is roughly $f_T/3$ for sufficient phase margin in the feedback loop. Furthermore, settling to >10-bit precision usually dictates about 10 time constants settling time. Together with the requirement that switched capacitor circuits need to settle in ½ clock period, we have

$$f_s \cong \frac{f_T}{3 \cdot 10 \cdot 2} \quad \frac{f_T}{60} \tag{3-23}$$

In literature, typical ratios of 25…100 have been stated [63, 64].
– Given the f_T requirement, we obtain the corresponding g_m/I_D from simulation data. I.e. for a given converter speed, we assume that the device is biased to yield f_T while maximizing g_m/I_D.
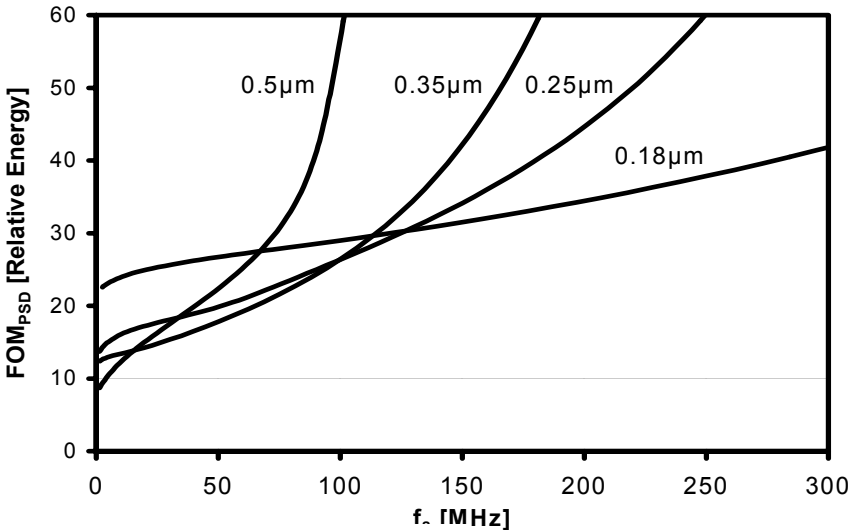
Figure 3-15. Noise limited circuit energy versus speed and technology.

The underlying equation for Figure 3-15 can be derived from (3-20) as

$$FOM_{PSD} \propto \frac{1}{V_{DD} \cdot \left(\dfrac{g_m}{I_D}\right)_{(f_T)}} . \tag{3-24}$$

From the result illustrated in Figure 3-15, we see that just like in the flash analysis, each technology becomes better than its predecessor only after a certain frequency boundary. However, the data in Figure 3-15 corrects the first order result that power should scale as $1/V_{DD}$. Depending on the speed requirements, technologies with lower V_{DD} may yield lower power.

5.2.4   Slewing

In the above discussion, we have assumed that the amplifier settles in a purely linear fashion. However, in practical switched capacitor circuits, the total settling time consists of a slewing and a linear settling time component. The total settling time is therefore

$$t_s \quad t_{slew} + t_{linear} . \tag{3-25}$$

It can be shown that the ratio between these two settling time components is approximately given by [65]

$$\frac{t_{slew}}{t_{linear}} \cong \frac{1}{N} \cdot \frac{g_m}{I_D} \cdot \left(\frac{V_{DD}}{G} - \frac{2}{g_m/I_D} \right), \tag{3-26}$$

where N is the number of required linear time constants and G is the closed loop gain of the amplifier. At first glance, (3-26) suggests that the slewing component should decrease for scaled technologies with small V_{DD}. This argument is also often found in previous literature, e.g. [2]. However, since smaller feature size devices tend to be operated at higher g_m/I_D, i.e. closer to "bipolar operation", the decrease in slewing time due to lower V_{DD} may be offset. To investigate this, we are showing the ratio t_{slew}/t_{lin} in Figure 3-16 for N=10 and G=2. A gain of G=2 is often used in pipeline stages to maximize their operating speed.
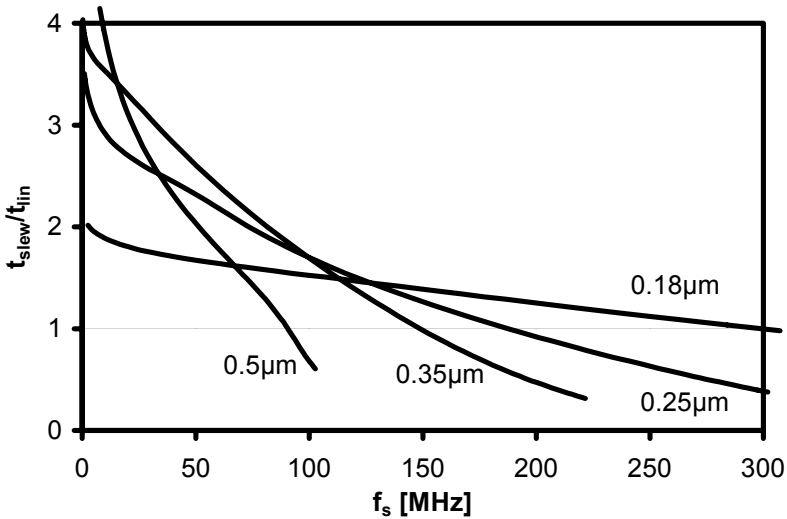
Figure 3-16. Ratio slewing/linear settling time vs. sampling speed.

As seen in Figure 3-16, the gain stage slews least in 0.18μm technology until about f_s=125MHz. After this point, the large required gate overdrive that is required to meet f_t in 0.25μm and 0.35μm technologies reduces the slewing component below that of 0.18μm.

It is now interesting to modify FOM_{PSD} to include the slewing effect. Using (3-25) and (3-26) we can rewrite the speed portion of (3-20) to get

$$FOM_{PSD} \propto \frac{1 + \dfrac{t_{slew}}{t_{lin}}}{V_{DD} \cdot \left(\dfrac{g_m}{I_D}\right)_{(fT)}} . \qquad (3\text{-}27)$$

We now plot this new figure of merit versus a new, effective sampling frequency that also captures the additional settling time due to slewing (see Figure 3-17). In Figure 3-17, the new, effective sampling rate is given by

$$f_s \quad \frac{f_T}{60 \cdot \left(1 + \dfrac{t_{slew}}{t_{lin}}\right)} \qquad (3\text{-}28)$$
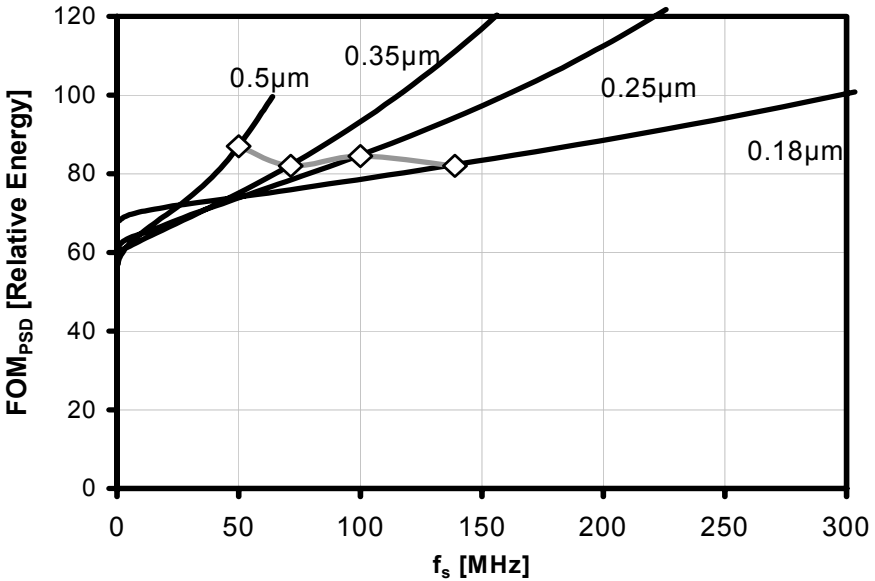
Figure 3-17. Noise limited circuit energy with slewing included.

Figure 3-17 indicates that that the energy efficiency as a function of technology becomes fairly flat when slewing is included. Yet, the data suggests that for a converter with f_s>50MHz implementation in the smallest feature size is most power efficient.

Just like in the analysis for matching limited power, we now construct a speed/scaling trajectory. For the four marked data points in Figure 3-17, we assume a typical average ADC speed of 50MHz in 0.5μm technology, and a throughput doubling every two process generations (see chapter 2). Again, this choice is somewhat arbitrary, but fairly reasonable. As we see from these four data points, energy efficiency in noise limited circuits is virtually constant, and independent of technology.

5.3   Pipelined ADC Performance Trends

We now compare the above result to *FOM*3 performance from published works. As an example, we use data from 10-bit pipelined ADCs, which is summarized in Table 3-2 and plotted in Figure 3-18 versus feature size. A linear fit to these data points shows an energy efficiency improvement rate

of roughly 2.2x over the 7.5 years. Equivalently, this corresponds to a 2x energy reduction every 6.6 years.

Table 3-2. 10-bit pipelined ADC performance.

Reference	Feature Size [μm]	Year	Speed [MS/s]	Power [mW]	Supply [V]	*FOM3*=Power/Speed [mW/MS/s]
[63]	0.8	1995	40	85	2.7	2.1
[66]	0.8	1999	40	119	3.3	3
[67]	0.8	1998	20	28	2.4	1.4
[68]	0.6	1998	14.3	36	1.5	2.5
[69]	0.6	1996	40	28	5	0.7
[70]	0.5	2001	200	280	3	1.4
[71]	0.35	2000	40	55	3	1.4
[72]	0.35	2000	100	105	3	1.1
[73]	0.35	1999	100	93	3	0.9
[74]	0.35	2000	10	15	3.3	1.5
[75]	0.3	2002	30	16	2	0.5
[76]	0.25	2000	20	43	1.4	2.2
[77]	0.25	1999	45	25	1.5	0.6
[78]	0.18	2001	80	80	1.8	1.0
[79]	0.18	2003	100	69	1.8	0.7
[80]	0.18	2003	150	100	1.8	0.7
[81]	0.12	2002	100	120	1.2	1.2
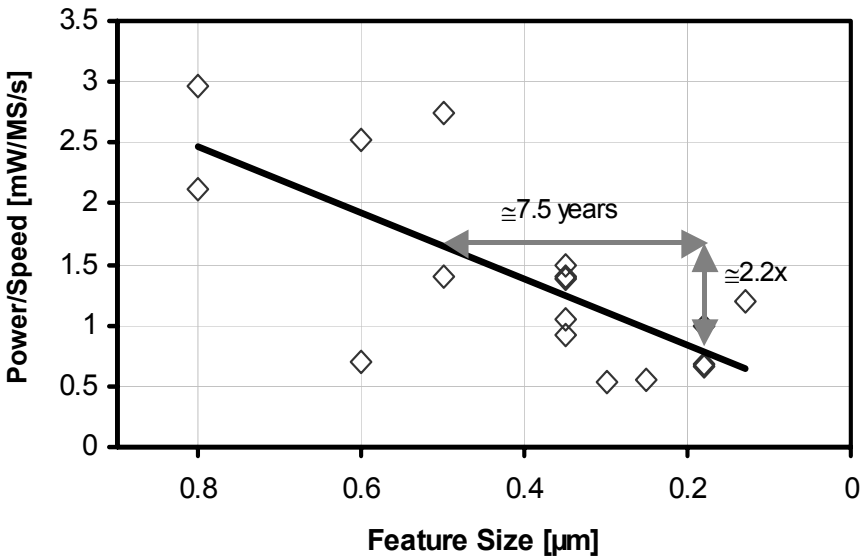
Figure 3-18. Published 10-bit pipelined ADC performance vs. technology.

5.4 Discussion

The above results show a discrepancy between the scaling behavior of noise limited transconductors and pipelined ADCs. While it is true that noise limited transconductors can dominate pipelined ADC power, we note that it may be quite inaccurate to extrapolate from a single transistor circuit to an entire A/D converter.

Pipelined converters extract a fixed number bits per stage. After resolving some of the bits in the front-end of the pipeline, the dynamic range of succeeding gain stages is usually scaled down to save power. As a result, only the first few stages of a pipeline ADC fall into the category of "noise limited transconductors" as analyzed above. Figure 3-19 shows a typical power distribution for a 10-bit pipeline [69].

In this design, only about 30% of the total power is dissipated in noise-limited amplifiers. Remaining parts of the pipeline consume "digital" power or "matching-limited" power, which decreases with further process scaling as shown in our analysis on flash ADCs. As a result, it becomes hard to quantify the scaling behavior exactly, unless the power dissipation profile of the converter implementation is considered.

Nevertheless, the analysis of this section corrects the viewpoint that the power in noise-limited circuits is bound to rapidly increase with technology scaling. The result suggests that noise-limited circuit efficiency remains constant with technology scaling. In addition, we saw that matching limited power, and also digital power consumption scale down with technology.
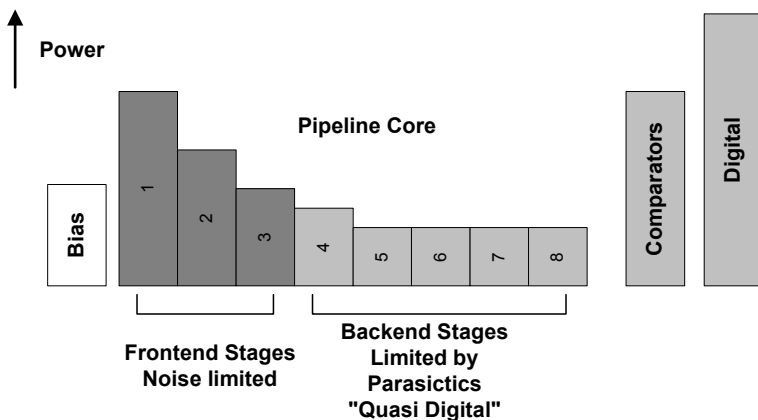
Figure 3-19. Typical 10-bit pipelined ADC power distribution.

This page intentionally left blank

Chapter 4

IMPROVING ANALOG CIRCUIT EFFICIENCY

1.   INTRODUCTION

As we have seen in chapter 2, power dissipation is one of the most severe showstoppers in the application of ADCs. This chapter aims to investigate opportunities for an increased level of "digital assistance" to help reduce the power dissipation, and potentially also increase the throughput of analog building blocks.

2.   ANALOG CIRCUIT CHALLENGES

Figure 4-1 below summarizes the main factors that determine analog circuit power dissipation. Fundamentally, the fairly low power efficiency in high performance analog signal processing originates from the simultaneous demand for high speed and precision.

While the power in analog circuits tends to grow linearly with the desired speed, the link to precision requirements is far more complex. From a general perspective, precision can be subdivided into three main components. The first and most fundamental limit in accuracy is given by the thermal noise of circuit elements. For example, the available signal headroom and the so-called "kT/C noise" [35] determine the dynamic range in an analog sampled data circuit. Reducing the standard deviation of the noise by a factor of two requires quadrupling the effective capacitance in the circuit. At constant speed, this necessitates a fourfold increase in transconductance, and hence a 4x increase in power dissipation.
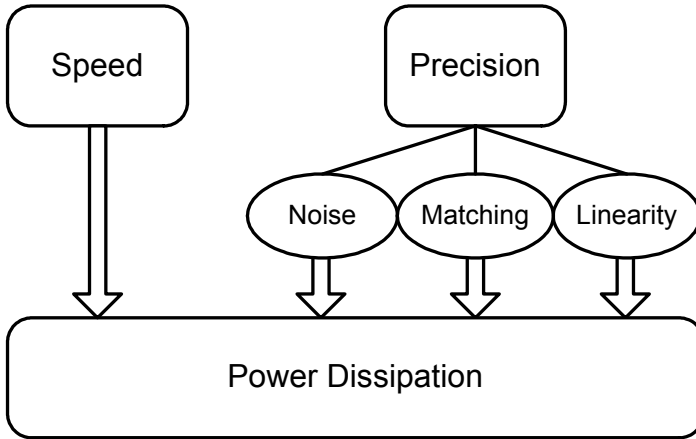
Figure 4-1. Analog circuit challenges and power dissipation.

In circuits that are limited by component matching, increasing the precision also translates into a power penalty. To first order, matching accuracy is inversely proportional to component area [38]. Therefore, additional precision requires larger components with larger capacitance and a resulting net increase in power dissipation. However, in contrast to thermal noise, matching errors are not fundamental, in the sense that they can be addressed without necessarily increasing component size. In many situations, it is possible to overcome matching errors using some form of trimming or calibration. In state-of-the-art ADCs digital correction and calibration techniques (e.g. [3],[4]) are routinely used to both avoid a matching-induced power penalty and to improve accuracy beyond technology limits.

A third significant challenge in precise analog signal processing arises from the need for highly linear amplification. In most electronic circuits, precisely linear operation is achieved by using high gain amplifiers in a negative feedback loop. In some sense, the use of electronic feedback parallels the approach of increasing component size to minimize mismatch. Achieving sufficient gain usually necessitates the use of complex amplifiers, with increased power dissipation and elevated noise. However, just as in the case of mismatch, distortion and gain inaccuracy limitations are not fundamental. Resulting errors can be compensated downstream, preferably through some digital compensation mechanism as well.

From this point of view, it is most interesting to investigate the potential advantage and power savings that are possible by lifting linearity

requirements in analog amplifiers. In the following section we examine the "cost of feedback" and the prospective power savings more closely.

3.   THE COST OF FEEDBACK

Figure 4-2 contrasts a typical precision amplifier (see e.g. [82]) with a simple open-loop gain stage. In the circuit of Figure 4-2(a), a high gain operational transconductance amplifier is used in a capacitive feedback configuration to achieve precise and drift insensitive voltage amplification. In principle, and leaving accuracy considerations aside, the simple resistively loaded differential pair of Figure 4-2(b) provides voltage amplification in an equivalent manner.

At first glance, it is clear that the cost for precision amplification is a much higher transistor count. As a result, there are also more noise contributors, which typically result in a significant power penalty. Aside from this obvious difference, a number of other aspects make the open-loop approach attractive, especially for an implementation in fine line technologies. First, the resistive loading effectively eliminates the need for high intrinsic transistor gain ($g_m \cdot r_o$), which is hard to achieve in transistors with short channel lengths (see section 3.4 in chapter 3).
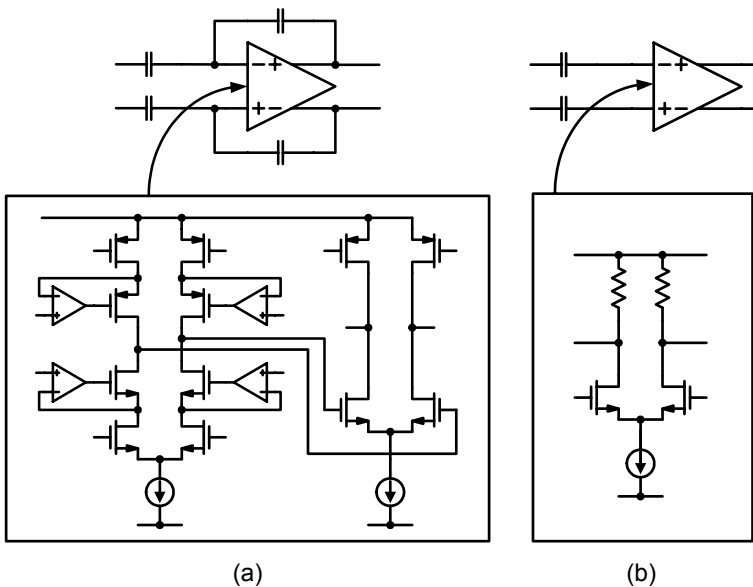
Figure 4-2. Comparison: (a) Precision feedback amplifier. (b) Open-loop amplifier.

Secondly, without the need for active loads, additional signal swing becomes available. This is most welcome in deep sub-micron technologies with diminishing headroom. A further advantage exists in the attainable bandwidth. In the feedback circuit, stability constraints limit the closed loop bandwidth of the amplifier to a fraction of the smallest non-dominant pole frequency. In the open-loop case, this constraint is removed.

Quantifying the precise net advantage of open-loop amplification in general is a challenging task. In the following discussion, we limit our analysis to the comparison of two reasonable and practical implementations and focus on the advantage in power efficiency.

4.   TWO-STAGE FEEDBACK AMPLIFIER VS. OPEN-LOOP GAIN STAGE

Consider the simplified, single ended circuits in Figure 4-3 for further analysis. The two-stage amplifier under consideration (Figure 4-3(a)) has become a standard topology in many ADCs, and is therefore useful for comparison. In this context, it is interesting to note that two-stage amplification has become necessary even in fairly low precision, 10-bit ADCs (e.g. [71, 80]). This is mainly because of the low intrinsic device gain in modern technologies and the low available headroom, which prevents the use of cascoded, telescopic amplifiers.
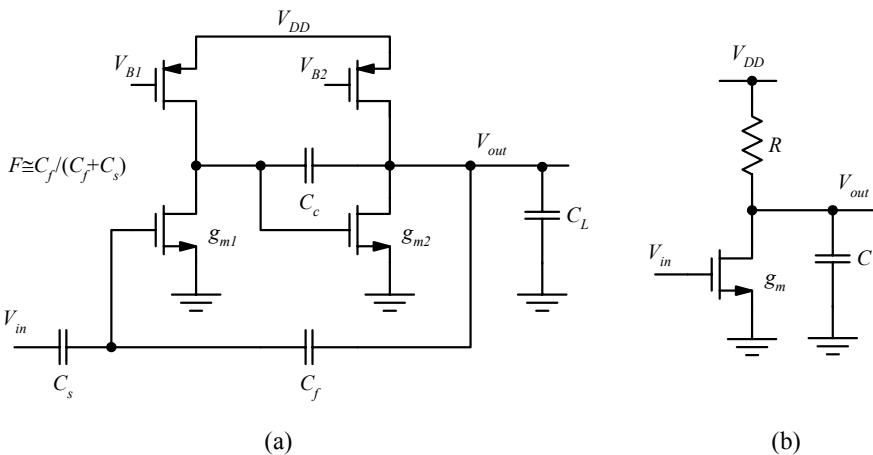
Figure 4-3. (a) Two-stage feedback amplifier. (b) Open-loop gain stage.

We now compare the two amplification approaches with respect to their power efficiency. A suitable figure of merit for this purpose is given by the FOM_{PSD} metric introduced in chapter 3 (see (3-21)). This figure of merit quantifies the amount of power that must be invested to obtain a certain speed and dynamic range.

Table 4-1 summarizes approximate, but sufficiently accurate expressions for the components of FOM_{PSD} for each amplifier (derivations are given e.g. in [65, 83]). The variables η in Table 4-1 represent the transconductor efficiency g_m/I_D in each circuit (see (3-3)).

Table 4-1. Amplifier performance metrics.

	Two-Stage Amplifier (Figure 4-3(a))	Open-Loop Amplifier (Figure 4-3(b))
Power	$V_{DD}(I_{D1}+I_{D2})$ $V_{DD}\cdot\dfrac{g_{m1}+g_{m2}}{\eta_a}$	$V_{DD}\cdot I_D$ $V_{DD}\cdot\dfrac{g_m}{\eta_b}$
Speed	$F\cdot\dfrac{g_{m1}}{C_c}$	$\dfrac{1}{RC}$
Dynamic Range	$\dfrac{\frac{1}{2}V_{ref}^2}{2\cdot\dfrac{1}{F}\cdot\dfrac{kT}{C_c}\cdot\left(1+F\dfrac{C_c}{C_L}\right)}$	$\dfrac{\frac{1}{2}V_{ref}^2}{(1+g_mR)\cdot\dfrac{kT}{C}}$

For simplicity in this comparison, we assume simple square law transistor models and equal gate overdrive $V_{OVa}=V_{GSa}-V_{TH}$ for all transistors in the feedback amplifier. Note that in an optimized design, the gate overdrives of the active loads may be chosen slightly larger to reduce the total noise of the amplifier. However, since the available supply headroom in modern technologies tends to prohibit this option, we will not consider it in this simplified discussion.

To simplify further, we neglect the effect of feedback network loading, both at the output and input node of the amplifier in Figure 4-3. The feedback factor F in this circuit is then related to the closed loop gain G of the gain stage by

$$F \quad \frac{1}{G+1}. \tag{4-1}$$

For a fair comparison, we assume that both circuits have an equivalent gain factor, i.e.

$$G \approx \frac{C_s}{C_f} \cdot \frac{1-F}{F} \approx g_m R.$$ (4-2)

Furthermore, we assume that a fixed reference voltage V_{ref} defines the peak output swing of each amplifier, rather than the supply limits. Note that with this assumption, we neglect the advantage of slightly larger available headroom in the open-loop topology. Within the accuracy of this analysis, this simplification has only minor impact on the final result. To include a suitable condition for stability in the feedback loop, we use

$$\frac{g_{m2}}{C_L} \approx 3 \cdot F \cdot \frac{g_{m1}}{C_c},$$ (4-3)

which corresponds to approximately 70 degrees phase margin. With these assumptions and simplifications, the FOM_{PSD} metric for each circuit becomes

$$FOM_{PSDa} \approx 4kT \cdot (1+G)^2 \frac{V_{DD}}{\eta_a \cdot V_{ref}^2}\left(1 + \frac{3}{G+1}\frac{C_L}{C_c}\right)\left(1 + \frac{1}{G+1}\frac{C_c}{C_L}\right)$$ (4-4)

$$\text{and } FOM_{PSDb} \approx 2kT \cdot G \cdot (1+G)\frac{V_{DD}}{\eta_b \cdot V_{ref}^2}.$$ (4-5)

Comparing (4-4) and (4-5), we note the following:
- The additional noise from active loads in the feedback amplifier results in a fundamental efficiency penalty (captured as a constant factor of 4 in (4-4) vs. 2 in (4-5))
- The two-stage feedback amplifier is subject to an additional, load dependent penalty.

To investigate the load dependence further, we plot the expression

$$\rho \approx \left(1 + \frac{3}{G+1}\frac{C_L}{C_c}\right)\left(1 + \frac{1}{G+1}\frac{C_c}{C_L}\right)$$ (4-6)

for several gain factors G and capacitor ratios C_c/C_L in Figure 4-4 below.
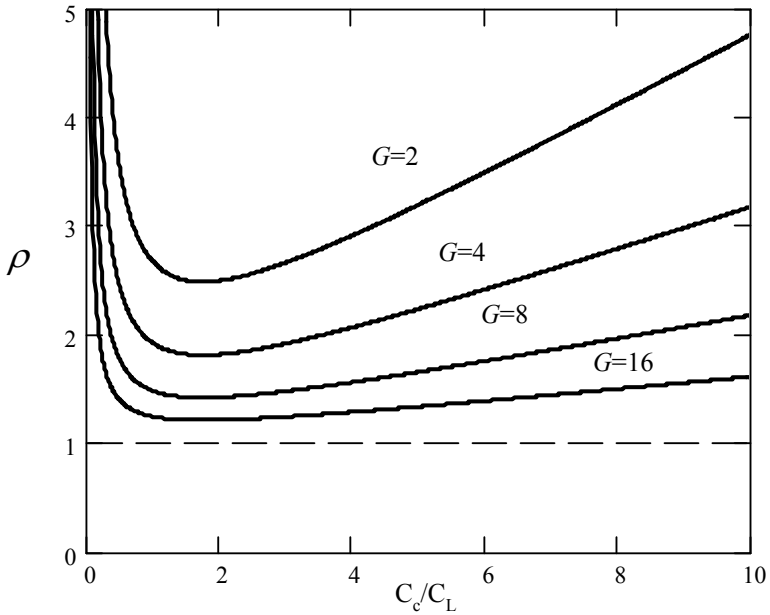
Figure 4-4. Two-stage amplifier penalty factor.

As apparent from these graphs, the load dependent penalty factor is a strong function of the desired closed loop gain *G*. Furthermore, for a fixed *G*, there exists a certain optimum value for C_c/C_L. To proceed with a conservative assumption, we assume that the feedback amplifier is always optimized for a minimum penalty ρ. This condition is met when

$$\frac{C_c}{C_L} \quad \sqrt{3} \, . \tag{4-7}$$

With this modification, (4-4) simplifies to

$$FOM_{PSDa} \quad 4kT \cdot (1+G)^2 \, \frac{V_{DD}}{\eta_a \cdot V_{ref}^2} \left(1 + \frac{\sqrt{3}}{G+1}\right)^2 \, . \tag{4-8}$$

We are now in the position to quantify the expected advantage of open-loop amplification directly. Using (4-5) and (4-8), the relative power savings are given by

$$S \quad \frac{FOM_{PSDa} - FOM_{PSDb}}{FOM_{PSDa}} \quad 1 - \frac{1}{2}\frac{\eta_a}{\eta_b}\frac{G}{1+G}\frac{1}{\left(1+\dfrac{\sqrt{3}}{G+1}\right)^2} . \tag{4-9}$$

Figure 4-5 plots this quantity for the case of $\eta_a = \eta_b$, i.e. equal transconductor efficiency and thus equal gate overdrive in both amplifiers. This result confirms the enormous potential for power savings when all linearity constraints are removed from the amplifying element. For any gain factor G, power savings of greater 60% seem possible.

There is, however, one additional factor that must be included in the final result. In a practical open loop amplifier, the "linear region" of the amplifying device(s) is limited. For instance, in a differential pair, complete current steering takes place when the input voltage exceeds $\sqrt{2} \cdot V_{OV}$, were V_{OV} is the quiescent point gate overdrive of each transistor [35]. Beyond this point, one differential pair transistor turns completely off and the amplifier gain drops to zero.
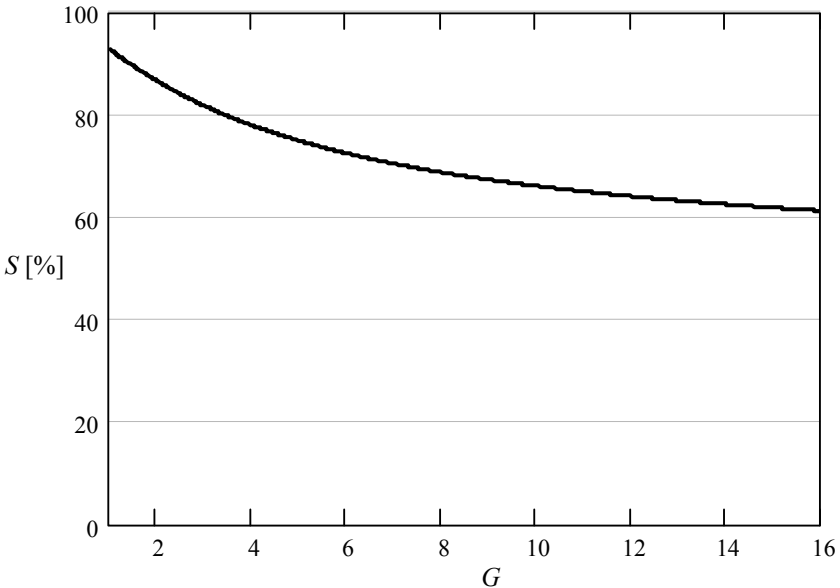
Figure 4-5. Percent power savings with open-loop amplification as a function of gain (assuming $\eta_a = \eta_b$).

To avoid this condition, we require

$$V_{OV} \geq \frac{1}{\sqrt{2}} V_{in\,max} \cong \frac{1}{\sqrt{2}} \frac{V_{ref}}{G} . \qquad (4\text{-}10)$$

Alternatively, the input range of the amplifier could be increased using resistive source degeneration. However, as discussed in the chapter 5, this modification reduces the efficiency of open-loop amplification and is therefore not considered in this analysis.

As we will also see in the next chapter, the limit case of equality in (4-10) is fairly impractical. For a reasonable compromise in the order of the introduced nonlinearity, it is reasonable to assume that the amplifier's input voltage can only span a fraction of the gate overdrive bias, i.e.

$$\alpha \cdot V_{OV} \quad V_{in\,max} \cong \frac{V_{ref}}{G} , \qquad (4\text{-}11)$$

with $\alpha < 1$. Since the transconductor efficiency in the open loop amplifier is inversely proportional to V_{OV}, this constraint imposes in some cases an upper bound on the achievable power savings. From the above, and using (3-3) we find

$$\eta_b \quad \frac{2 \cdot G \cdot \alpha}{V_{ref}} . \qquad (4\text{-}12)$$

With this constraint, we now construct a modified version of Figure 4-5 with typical design values of $V_{ref}=1V$ and η_a set to a reasonably practical maximum value of $\eta_{max}=10V^{-1}$ ($V_{OVmin}=200mV$) for a high-speed design. Since this upper bound also applies to η_b, we have

$$\eta_b \quad \min\left(\frac{2 \cdot G \cdot \alpha}{V_{ref}}, \eta_{max} \right). \qquad (4\text{-}13)$$

Figure 4-6 shows the resulting power savings plot for several values of fractional swing α. As we can see, the swing constraint severely limits the open-loop advantage for small gains G and low choices for α. Nevertheless, the anticipated power advantage from open loop amplification is significant, under virtually any design condition.
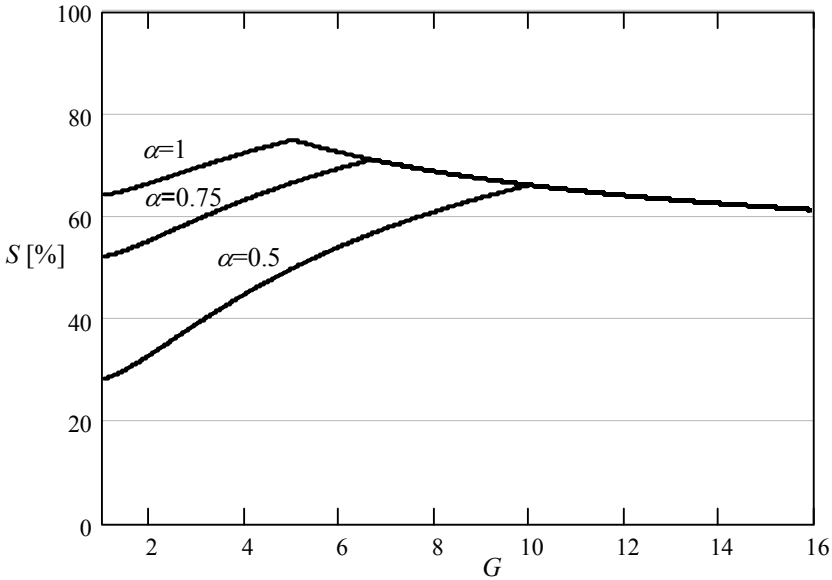
Figure 4-6. Percent power savings with open-loop amplification as a function of gain (assuming $V_{ref}=1V$, $\eta_a=10V^{-1}$ and η_b given by (4-13)).

5.   DISCUSSION

Despite the large number of simplifications made in the preceding analysis, the projected advantage of Figure 4-6 matches the outcome of the prototype design described in later chapters closely. With respect to future technology scaling, the limitation due to input swing constraints is expected to decrease in magnitude. This can be seen from equation (4-13). Decreasing supply voltages may necessitate a drop in the reference voltage V_{ref}, which leads to a more advantageous lower bound for the open-loop transconductor efficiency.

In the following chapters, we describe in more detail how the proposed open-loop amplification concept can be used to its full advantage for the specific case of a pipelined ADC implementation. The discussion will begin by developing analog circuit design guidelines, followed by a derivation of the required digital nonlinearity correction mechanism.

Chapter 5

OPEN-LOOP PIPELINED ADCS

1. A BRIEF REVIEW OF PIPELINED ADCS

Pipelined converters have become the predominant architecture for ADCs with resolutions of 8-14 bits and conversion rates of 10-200MS/s. Figure 5-1 shows a conceptual block diagram of this converter topology. Several converter stages are cascaded and process the analog input sequentially, analogous to flip-flops propagating a bit stream in a digital shift register.
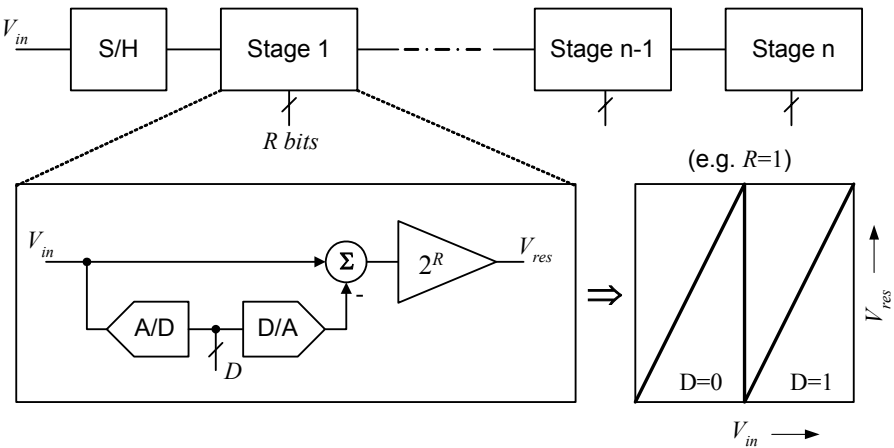
Figure 5-1. Pipelined ADC block diagram.

Each stage performs a sample and hold operation and a coarse A/D conversion. The local quantization result is converted back into analog form and used to compute the error in the coarse digital approximation D. The locally computed and amplified quantization error, often called the residuum (V_{res}), propagates through subsequent stages which resolve further less significant digital information of the initial input sample. After the signal has passed through all stages, the sub quantization results are combined to yield the final digital output word.

The main advantage of this architecture is that due to stage pipelining, its throughput rate is set by the time needed to perform a single sub-A/D and D/A conversion. The fact that the signal needs to propagate though all stages until the final conversion result becomes available results only in conversion latency, which is tolerable in many signal processing applications.

Also shown in Figure 5-1 is an ideal pipeline stage transfer function, V_{res} as a function of stage input voltage V_{in}, for the simple case of a 1-bit sub-conversion. In this example, the residuum segments have a slope of 2. More generally, it can be shown that in a stage that resolves R bits, a gain factor of 2^R is needed.

2.   CONVENTIONAL STAGE IMPLEMENTATION

Figure 5-2 shows a conceptual, single ended block diagram of a conventional pipeline stage. This circuit consists of a flash-type sub-ADC, a capacitive charge redistribution network, and a high performance transconductance amplifier (G_m block in Figure 5-2).

The stage operates in two main clock phases. During the sampling phase, the input signal V_{in} is acquired. In a second phase, a residual charge packet, controlled by the local conversion result D, is redistributed onto the feedback capacitor C_F to produce the amplified stage residuum V_{res}. In this conventional scheme, the use of electronic feedback around the amplifier results in a precise and drift insensitive stage transfer function. However, as discussed in the previous chapter, the cost of this desirable feature is an excessive voltage gain requirement. In the front-end of high-resolution pipelines (e.g. [82]), two-stage, gain boosted amplifiers with open-loop gain >100dB are often needed to meet the stringent accuracy requirements.

Since the precision of all other components in a pipelined converter can be relaxed using existing digital correction techniques [3, 4], residue amplifiers dominate the overall power dissipation. A contribution of up to 50-70% to the total ADC power is typical.
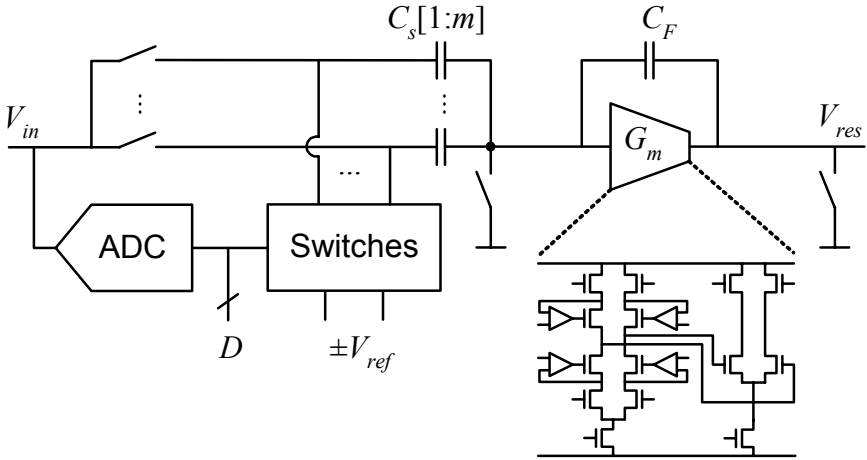
Figure 5-2. Conventional pipeline stage.

As a result, a variety of techniques have been developed to minimize amplifier power in pipelined ADCs. Among them, stage scaling [84], [85], optimization of the per-stage-resolution [86-88], and amplifier sharing techniques [79, 89, 90], are commonly used.

In addition to their dominance in power consumption, it has also been recognized that residue amplifiers are most susceptible to complications that arise from continuing integrated circuit technology scaling [27], [25]. For implementations in future deep sub-micron processes, it is often predicted that limited supply headroom and low intrinsic device gain may lead to a relative power increase in such noise-limited, precision analog circuit blocks [30], [29].

Replacing precision residue amplifiers with simple open-loop stages and correcting for the resulting errors digitally is a solution that helps mitigate both of the above-mentioned issues. In the following section we discuss basic design considerations for the proposed open-loop pipeline stages.

3. OPEN-LOOP PIPELINE STAGES

Recently, the benefits of using open-loop structures in high-speed pipelined ADCs have been recognized and demonstrated. The 8-bit ADCs reported in [13] and [91] use open-loop, current mode residue amplification to achieve excellent power efficiency at high conversion speeds. In this book, we propose a voltage mode topology in conjunction with digital calibration to extend the applicability of open-loop structures to resolutions

of greater 10 bits. Figure 5-3 shows a conceptual schematic diagram of the proposed stage implementation.

Except for the charge redistribution phase, the operation of this circuit is similar to the conventional topology described above. Unlike in the closed-loop implementation, the residual charge packet on the capacitive array is not redistributed onto a feedback capacitor, but remains in place to produce a small voltage at node V_x. This residuum is fed into a resistively loaded transconductance stage to produce the desired full-swing residue voltage V_{res}. Since the high gain requirement in the transconductor is now dropped, a simple differential pair can be used to replace the complex amplifier in Figure 5-2. As we have argued in the previous chapter, this modification results in significant power savings and also mitigates technology-scaling issues. These advantages, however, come at the price of several new non-idealities in the stage transfer function that have not been addressed in previous work.

3.1   Open-Loop Stage Analysis

With sufficient loop gain in the conventional implementation of Figure 5-2, deviations of the stage transfer function from ideality are mostly due to capacitor mismatch and offset errors in the coarse sub-ADC. With the introduction of the simplified, open-loop amplifier of Figure 5-3, several additional error sources must be considered. Figure 5-4 depicts an appropriate model for further analysis.
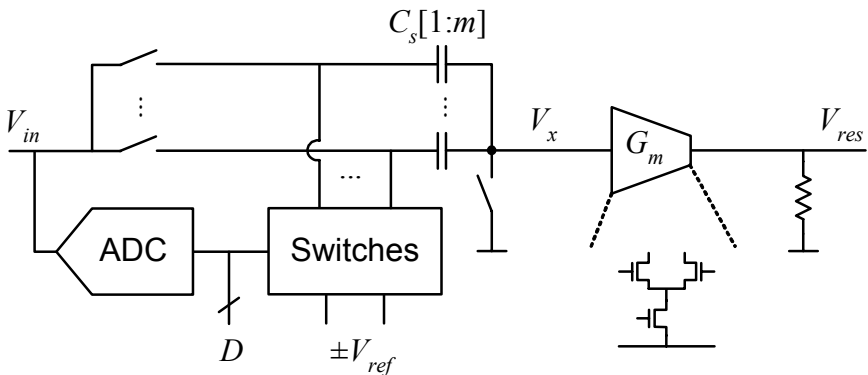
Figure 5-3. Open-loop pipeline stage.
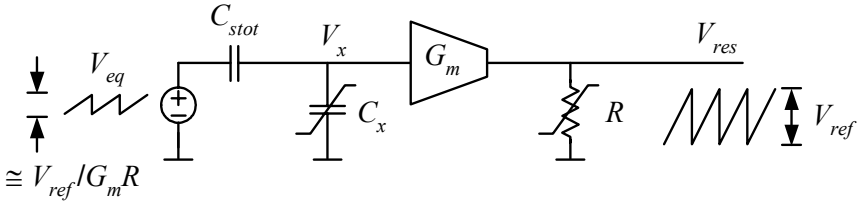
Figure 5-4. Open-loop stage model.

Here, the capacitor array is replaced with its Thévenin equivalent, consisting of the total array capacitance C_{stot} and an equivalent voltage source V_{eq} that represents the local stage residuum before amplification. Ideally, the transfer function from V_{eq} to the output V_{res} should be linear with a precise gain of 2^R, where R is the effective stage resolution. In the circuit of Figure 5-4, the transfer function is neither linear nor precisely defined. The linear gain term from source to output is set by the amount of parasitic capacitive attenuation at node V_x and the $G_m{\cdot}R$ product, which typically cannot be accurately controlled.

Furthermore, the amplification is nonlinear, primarily due to three effects: (1) voltage dependence of the capacitor C_x, which represents the transconductor input capacitance and parasitic junctions, (2) nonlinearity in the resistive load, and (3) nonlinearity in the *V-I* relationship of the transconductor. With respect to the tolerable errors in a pipelined ADC, none of the above nonlinearities may be negligible. However, for a practical and optimized implementation, it is reasonable to assume that the differential pair dominates the overall cascade nonlinearity that links V_{eq} and V_{res}. In the following analysis, we therefore focus on this particular error component, noting that some additional, but non-dominant distortion is actually due to other non-idealities. A more detailed derivation of nonlinear effects in the open-loop charge redistribution is presented in the appendix.

3.2   Distortion Model

Distortion in semiconductor devices can be partitioned into static and dynamic components. Dynamic, frequency dependent distortion is usually present when a nonlinear circuit is operated near its pole frequencies, in which case memory effects become significant [92, 93].

The class of switched capacitor circuits considered here is typically designed such that all node voltages settle to within a small fraction of an

LSB to their final "DC" values, i.e. the asymptotic values for infinite settling time. In this case a simple memoryless model, based on a power series of the form

$$y \quad a_1 x + a_2 x^2 + a_3 x^3 + ... \tag{5-1}$$

is sufficient for further analysis. Assuming ideal square law transistor models, one can express the differential pair *V-I* relationship as [92]

$$\frac{\Delta I}{I_{SS}} \quad \left(\frac{V_x}{V_{OV}}\right) + \frac{1}{4}\frac{\Delta\beta}{\beta}\left(\frac{V_x}{V_{OV}}\right)^2 - \frac{1}{8}\left(\frac{V_x}{V_{OV}}\right)^3 - \frac{1}{128}\left(\frac{V_x}{V_{OV}}\right)^5 - ..., \tag{5-2}$$

where ΔI and I_{SS} are the differential pair output and tail current respectively, V_{OV} is the quiescent point gate overdrive (V_{GS}-V_{TH}), and $\Delta\beta/\beta$ is the current factor mismatch of the two transistors. This transfer function is illustrated graphically in Figure 5-5.

Figure 5-6 shows the relative peak magnitude of the nonlinear terms in equation (5-2) as a function of the fractional input swing α, which is given by

$$\alpha \quad \frac{V_{x\max}}{V_{OV}}, \tag{5-3}$$
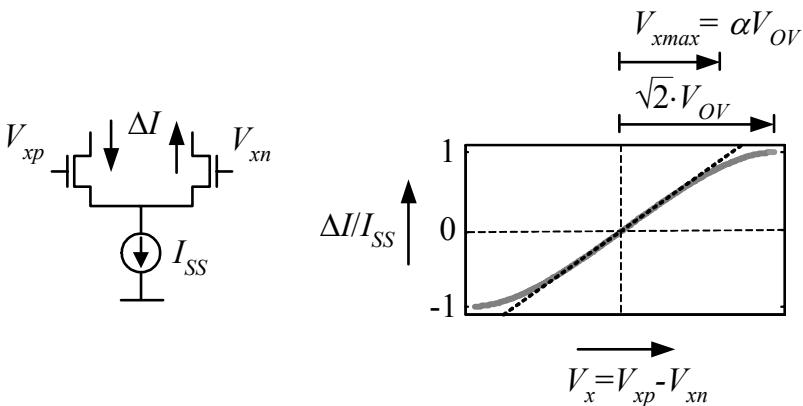
Figure 5-5. Differential pair *V-I* characteristic.
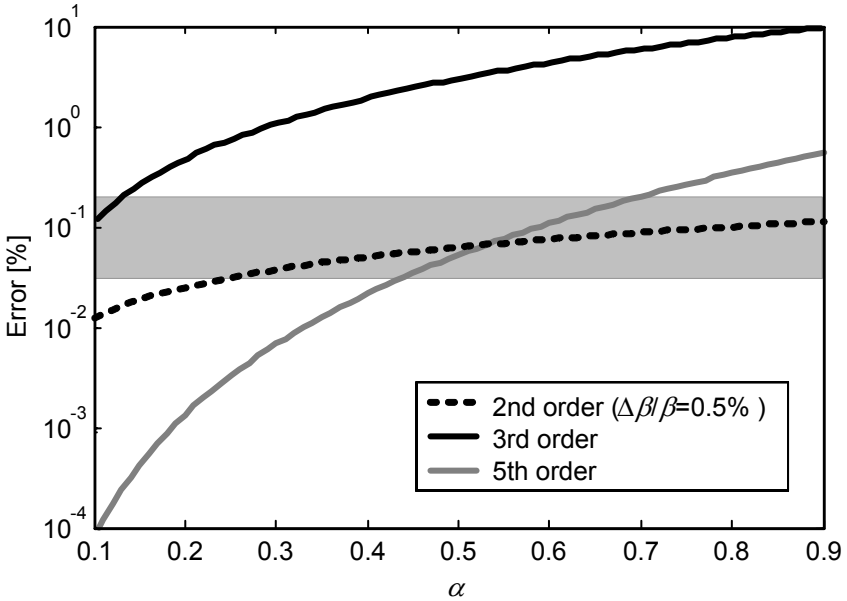
Figure 5-6. Differential pair nonlinearity as a function of $\alpha = V_{xmax}/V_{OV}$.

In equation (5-3), V_{xmax} is the peak magnitude of the differential pair's input voltage (see also chapter 4). V_{xmax} is usually fixed and determined by the chosen stage gain G and converter reference voltage V_{ref} since

$$V_{x\max} \cong \frac{V_{ref}}{G}. \qquad (5\text{-}4)$$

For the second order component in Figure 5-6 we assume a transistor matching of $\Delta\beta/\beta$ 0.5%, which is a typical achievable value in current CMOS technology. The shaded area is the range of typical precision requirements in pipelined ADCs. For resolutions of 8-14 bits, amplification to half-LSB precision translates into tolerable errors on the order of 2^{-9}-2^{-15} or roughly 0.2-0.003%.

As is apparent from Figure 5-6, choosing small fractional swing α, or equivalently, a large gate overdrive V_{OV} results in small nonlinearity errors. However, as shown in chapter 4, this may translate into a power penalty, especially for small stage gains. In principle, if all nonlinearity components including 5th and higher order distortion errors are removed in the digital domain, the choice of α is not critical and can approach one.

In order to achieve a reasonable compromise between analog power savings and digital post-processing complexity, we focused in this work on

compensating only nonlinearity components up to 3rd order. Under this circumstance, α must be chosen small enough to make 5th and higher order terms negligible. For instance, keeping the 5th order error below 0.1% translates into an upper bound for α of approximately 0.6.

In the context of the above discussion, it should be noted that the expression given in equation (5-2) tends to overestimate the distortion for short channel transistors with velocity saturation. In principle, velocity saturated transistors can be modeled as weakly degenerated square law devices [35], which leads to a reduction in the expected nonlinearity. For a given technology, more precise values for the coefficients in (5-2) may be obtained using simulations with appropriate short channel transistor models. However, for basic design considerations, (5-2) can be regarded as a conservative, sufficiently accurate expression. Furthermore, the digital compensation approach described in the following chapters adapts to each distortion component individually and does not assume a precise relationship between the associated coefficients.

4. ALTERNATIVE TRANSCONDUCTOR IMPLEMENTATIONS

Several alternatives exist for the implementation of the G_m-stage in Figure 5-4. As we have seen, one critical aspect in the design and efficiency of the transconductor is its useable input range.

One way to extend the input range of a differential pair is to use resistive source degeneration [35]. With this approach, the allowable input swing increases by $(1+T)$, where $T=g_mR_s$ is the local loop gain introduced by the degeneration. However, since the local feedback also reduces the compound transconductance by the same factor, there is, to first order, no net gain in the efficiency of the transconductor.

On the other hand, degeneration may be useful when implementing very small stage gains. A gain of one or two may require a gate overdrive on the order of 1V or larger, which typically translates to very small, poorly matched devices and also a severe penalty in the available signal headroom. In such cases, source degeneration in a folded topology as used in [94] should be considered.

Another potential advantage for resistive source degeneration comes from a noise perspective. Analysis shows that to first order, local feedback has little impact on the noise performance of a differential pair. However, for very short channel transistors with excessive thermal noise [44], degeneration with resistors of lower power spectral noise density can be advantageous [95].

Aside from the above-discussed possibilities, there are a number of other techniques that haven't found widespread use, but could be re-visited in the context of this work. Among them are the use of a g_m-boosted degeneration approach [96], addition of a gain expansive parallel circuit [97] and the combination of several offset differential pairs [98].

This page intentionally left blank

Chapter 6

DIGITAL NONLINEARITY CORRECTION

1.   OVERVIEW

This chapter describes a purely digital post-processing circuit that is capable of correcting errors from imprecise open-loop amplification in pipeline stages.

Figure 6-1(a) shows the general architecture of an n-stage pipelined converter with the proposed compensation scheme. The overall structure is canonical, with dedicated calibration circuitry added on a per-stage basis. The required digital hardware consists of three main components. The blocks labeled "correction" consist of digital arithmetic that is used to compensate for analog domain non-idealities. The remainder of this chapter focuses on details of this functional block.

A second set of blocks, labeled "estimation", assumes the task of identifying optimal correction parameters, based on the system's response to a binary random number generator (RNG) modulation sequence. The details of this technique are described in chapter 7.

As depicted in Figure 6-1(a), we assume that several uncritical converter stages in the pipeline backend do not require calibration. The calibration process of the more critical front-end stages is nested and shows similarities to the "accuracy bootstrapping" approach of [4] and [99]. Conceptually, the errors of the first, least significant stage under calibration (stage i in Figure 6-1) are measured and corrected using the succeeding, sufficiently accurate stages. Once the correction parameters of stage i have been determined, the algorithm proceeds with the calibration of stage i-1 and works its way toward the front-end of the converter.
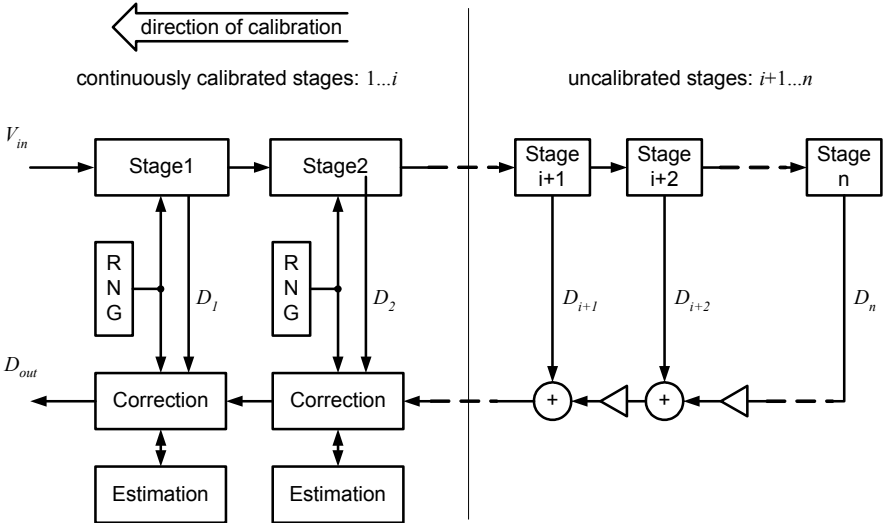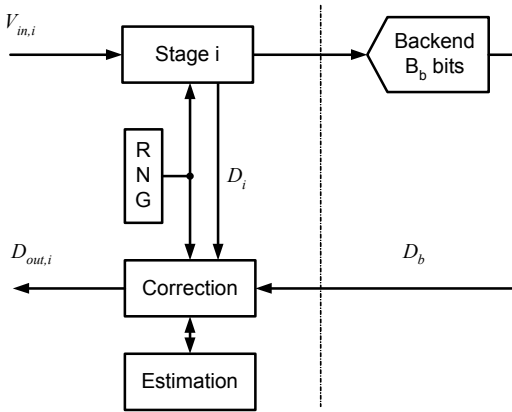
Figure 6-1. (a) ADC block diagram. (b) Reduced model for analysis.

In the particular implementation of the algorithm as a background calibration technique (see chapter 7), all correction parameters are continuously estimated and updated in a concurrent rather than sequential fashion. Upon startup of the system, however, parameter convergence occurs from stage *i* toward the front-end as explained above.

For simplicity in this analysis, we focus on the compensation of the least significant, i^{th} converter stage only. In the reduced system of Figure 6-1(b), the backend stages $i+1…n$ are modeled as an ideal ADC with an effective resolution of B_b bits. An extension of the obtained results for multi-stage calibration is straightforward and shows similarity to previously published analyses [99]. The prototype implementation that is described in chapters 8 and 9 of this book corresponds to the degenerate case with $i=1$, i.e. only the first converter stage is being calibrated.

2. ERROR MODEL AND DIGITAL CORRECTION

In order to establish a model for the required hardware in the digital correction blocks, we represent the front-end pipeline stage as a sub-circuit that consists of a coarse sub-ADC and sub-DAC, a differencing node and an interstage gain element. The resulting system model is shown in Figure 6-2. For notational convenience, we drop the stage index i from all variables and consider both analog and digital signals as unitless quantities whose full scale ranges are normalized to one.

Within the scope of this analysis, we also assume that both the sub-DAC and the differencing node are ideal, and only the amplifier and sub-ADC deviate from their ideal characteristics. In principle, the calibration concept could be extended to correct for DAC non-idealities as well (see section 8 of chapter 7).
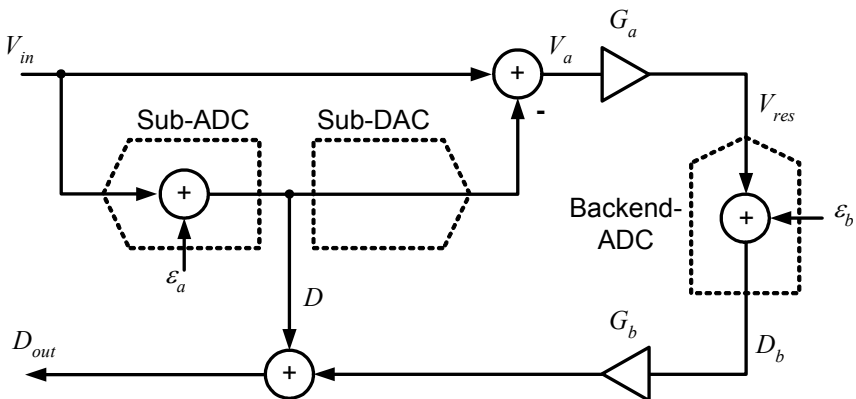

Figure 6-2. Reduced model with stage sub-circuits.

2.1   **Linear Amplifier Model**

For simplicity, we first consider the case of a perfectly linear interstage gain element. For this particular case, the required digital correction arithmetic consists only of a summing element that combines the coarse sub-ADC result D with an appropriately scaled version of the backend data D_b (see e.g. [28]). Writing the transfer function from the analog input to the digital output of the system in Figure 6-2, we obtain

$$D_{out} \quad V_{in} + \varepsilon_a\left(1 - G_a G_b\right) + \varepsilon_b G_b , \qquad (6\text{-}1)$$

where G_a and G_b are the linear gain factors in the analog and digital signal paths, and ε_a and ε_b represent quantization noise and other error sources in the coarse sub-ADC and the pipeline backend respectively. For the case of perfect weighting, i.e. $G_b = 1/G_a$, equation (6-1) reduces to

$$D_{out} \quad V_{in} + \frac{\varepsilon_b}{G_a} \qquad (6\text{-}2)$$

This expression corresponds to perfect operation of the pipelined ADC and implicitly assumes that the stage residue $V_{res} = -G_a\varepsilon_a$ does not exceed the full-scale input range of the backend converter. Note that in practice, this can be ensured by introducing redundancy through either one or a combination of three techniques: (a) using a "reduced radix" for G_a [4], (b) addition of redundant sub-ADC comparators to reduce ε_a [3], or (c), addition of redundant comparators in the backend converter to expand its full scale input range [99].

Assuming sufficient redundancy, equation (6-2) confirms the well known result that the overall conversion error is independent of sub-ADC errors and simply given by the backend quantization error reduced by the gain of the preceding analog signal path. This fact is an important prerequisite for the description of the pseudo-random modulation in chapter 7, and also insightful for comparison with the more general nonlinear amplifier compensation scheme described below.

2.2   **General Nonlinear Amplifier Model**

Consider now the case in which the gain element of the pipeline stage has an arbitrary, nonlinear transfer function of the form $V_{res} = g_a(V_a)$. Replacing the linear relationships $G_a \cdot V_a$ and $G_b \cdot D_b$ of Figure 6-2 with the general functions $g_a(V_a)$ and $g_b(D_b)$ modifies equation (6-1) to

$$D_{out} \quad V_{in} + \varepsilon_a + g_b[g_a(-\varepsilon_a) + \varepsilon_b]. \tag{6-3}$$

Inspecting this equation, we see that the precise cancellation of the sub-ADC error ε_a requires knowledge of the backend quantization error ε_b. However, under the assumption that g_b is only weakly nonlinear over the range of the small additive term ε_b, we can use a first order Taylor expansion to approximate (6-3) with

$$D_{out} \cong V_{in} + \varepsilon_a + g_b[g_a(-\varepsilon_a)] + \varepsilon_b \left. \frac{dg_b}{dD_b} \right|_{D_b \ g_a(-\varepsilon_a)}. \tag{6-4}$$

Provided that g_a is invertible, we can now choose the digital weighting function as the inverse of the amplifier transfer function, i.e. $g_b = g_a^{-1}$, to obtain

$$D_{out} \cong V_{in} + \varepsilon_b \left. \frac{dg_b}{dD_b} \right|_{D_b \ g_a(-\varepsilon_a)} \quad V_{in} + \varepsilon_b \left(\left. \frac{dg_a}{dV_a} \right|_{V_a \ -\varepsilon_a} \right)^{-1}. \tag{6-5}$$

Just as in the linear case, this ideal choice of digital weighting results in perfect cancellation of the sub-ADC error ε_a. In the limit case of a perfectly linear interstage gain element, the residual quantization error term in (6-5) reduces to the form of (6-2), with a constant term dividing the backend error ε_b.

With nonlinearities present, (6-5) predicts a signal dependent and hence nonlinear deviation from the ideal case. With respect to the staircase transfer function of the overall ADC, it is straightforward to show that this modulation causes systematic positive or negative differential nonlinearity (DNL) in regions with decreasing or increasing slopes of g_a respectively. However, for a reasonable and practical interstage amplifier design with only moderate nonlinearity, the expected penalty without further correction is low compared to a typical DNL budget and consequently not addressed in this work. For instance, a slope change of 10% over the full-scale range of amplifier's transfer function will only result in a DNL of approximately 0.1LSB.

A more stringent limitation to the attainable precision in the nonlinear error correction stems from practical considerations. Consider the more generalized error correction model in Figure 6-3 for further discussion.
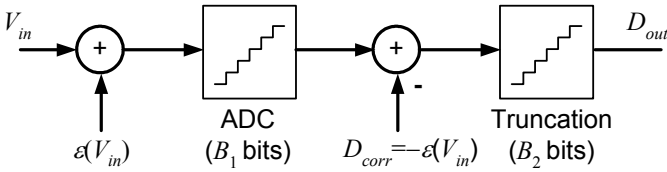
Figure 6-3. Model for error compensation.

This general setup shows an ADC with its input referred conversion error, and digital domain compensation. Even though the digital domain correction word D_{corr} can be made arbitrarily precise by choosing a large bit width, the final converter result must be truncated to a width that is suitable for the application. The truncation corresponds to a re-quantization, and thus, the overall quantization error is given by the series effect of the ADC quantization and subsequent truncation.

Assuming that ε is a linear function of V_{in}, and $B_1=B_2=B$, it can be shown that the total quantization error is upper bounded by ½ LSB at the B-bit level. If ε is a nonlinear function of V_{in}, the quantization levels of the truncation are projected onto the levels of the ADC in a distorted fashion. In this case, the upper bound of the total quantization error becomes 1 full LSB at the B-bit level, and hence ½ LSB excess conversion error compared to an ideal ADC.

To remedy this problem, it can be shown that either B_1 or B_2 must be increased beyond the desired ADC resolution. In the implementation of chapter 8, we chose $B_1=B_2+2$. This yields an upper bound for the quantization error of ½ LSB + ¼·½ LSB = 5/8 LSB, or equivalently a maximum ADC error penalty of one-eighth LSB. Note that adding these extra bits can be achieved simply by adding extra stages to the backend of the pipeline. The addition of redundant stages for calibration purposes is common to most pipelined ADC calibration principles, and is known to cause little power and area overhead.

2.3   Polynomial Amplifier Model

As we have shown in the previous chapter, we can achieve a compact third order nonlinearity model by appropriately choosing the gate overdrive of the open-loop amplifier. With the symbol conventions used in this chapter we then have

$$g_a(V_a) \quad V_{res} \quad a_1V_a + a_2V_a^2 + a_3V_a^3 . \tag{6-6}$$

Note that in this expression, the ideal value of a_1 is given by 2^R, where R is the desired effective stage resolution in bits. As shown above, the digital domain nonlinearity cancellation is accomplished by computing the inverse of the residue amplifier transfer function. Even though an explicit form for the inverse of (6-3) exists, a more efficient approach for the weakly nonlinear systems considered here is to perform the correction through additive terms that capture only the small deviation from ideality. This approach is illustrated in Figure 6-4. The correction function $e(D_b)$ operates on the raw backend data to cancel nonlinear components in the amplifier transfer function. Subsequently, an appropriate linear scaling operation is needed before assembling the final digital output.

The required compensation function $e(D_b)$ can be found by first expressing the nonlinear error in terms of the amplifier input V_a as

$$e(V_a) \quad g_a(V_a) - a_1V_a . \tag{6-7}$$

To write this error as a function of the amplifier output and ultimately as a function of the digital backend code D_b, we use the fact that $V_a = g_a^{-1}(V_{res})$ to obtain

$$e(V_{res}) \quad g_a\left(g_a^{-1}(V_{res})\right) - a_1g_a^{-1}(V_{res}) \quad V_{res} - a_1g_a^{-1}(V_{res}) . \tag{6-8}$$
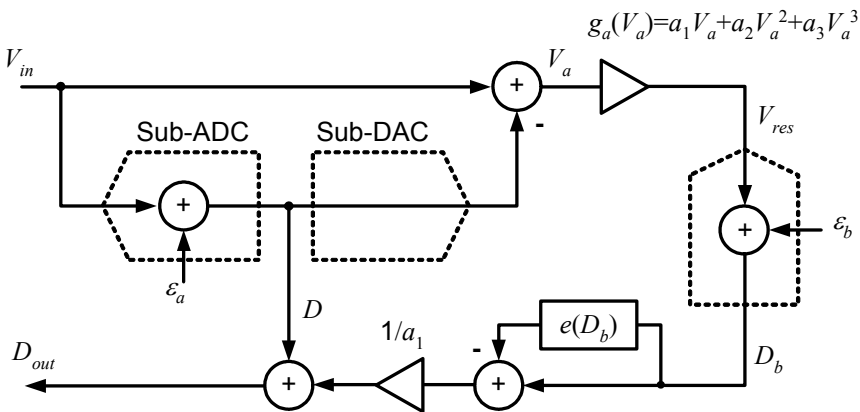

Figure 6-4. Additive nonlinearity compensation.

Substitution of $V_{res}=D_b-\varepsilon_b$ and making appropriate approximations yields

$$e(D_b) \cong D_b - a_1 g_a^{-1}(D_b) - \varepsilon_b \left[1 - a_1 \left(\frac{dg_a}{dV_a} \bigg|_{V_a - \varepsilon_a} \right)^{-1} \right]. \tag{6-9}$$

By the same argument as in section 2.2 above, the last term in this expression is small in a weakly nonlinear system and can be disregarded in further analysis. We now take further steps to transform (6-9) into a closed form expression that relates $e(D_b)$ directly to the polynomial coefficients of (6-6). For finding the inverse $g_a^{-1}(D_b)$, it is advantageous to note that

$$g_a(V_a) \quad a_1 V_a + a_2 V_a^2 + a_3 V_a^3 \quad b_0 + b_1(V_a - V_s) + b_3(V_a - V_s)^3 \tag{6-10}$$

with

$$b_0 \quad \frac{2a_2^3}{27a_3^2} - \frac{a_1 a_2}{3a_3}$$

$$b_1 \quad a_1 - \frac{a_2^2}{3a_3}$$

$$b_3 \quad a_3 \tag{6-11}$$

$$V_s \quad -\frac{a_2}{3a_3}$$

These equalities hold under the condition that $a_3 \neq 0$. In practice, assuming a fully differential circuit implementation, this condition is not restrictive since cubic nonlinearity is usually the dominant distortion mechanism. With the definition of shifted variables $V_{a1}=V_a-V_s$ and $V_{res1}=V_{res}-b_0$, we can rewrite (6-10) as

$$g_{a1}(V_{a1}) \quad V_{res1} \quad b_1 V_{a1} + b_3 V_{a1}^3. \tag{6-12}$$

As we will see shortly, this substitution greatly simplifies the complexity and number of parameters needed in the inverse function contained in (6-9). However, the applied coordinate shifts must be considered and compensated through some alternative mechanism in the system. Consider Figure 6-5(a) for further investigation.

$$g_{a1}(V_{a1}) = b_1 V_{a1} + b_3 V_{a1}{}^3$$

$$g_{a1}(V_{a1}) = b_1 V_{a1} + b_3 V_{a1}{}^3$$
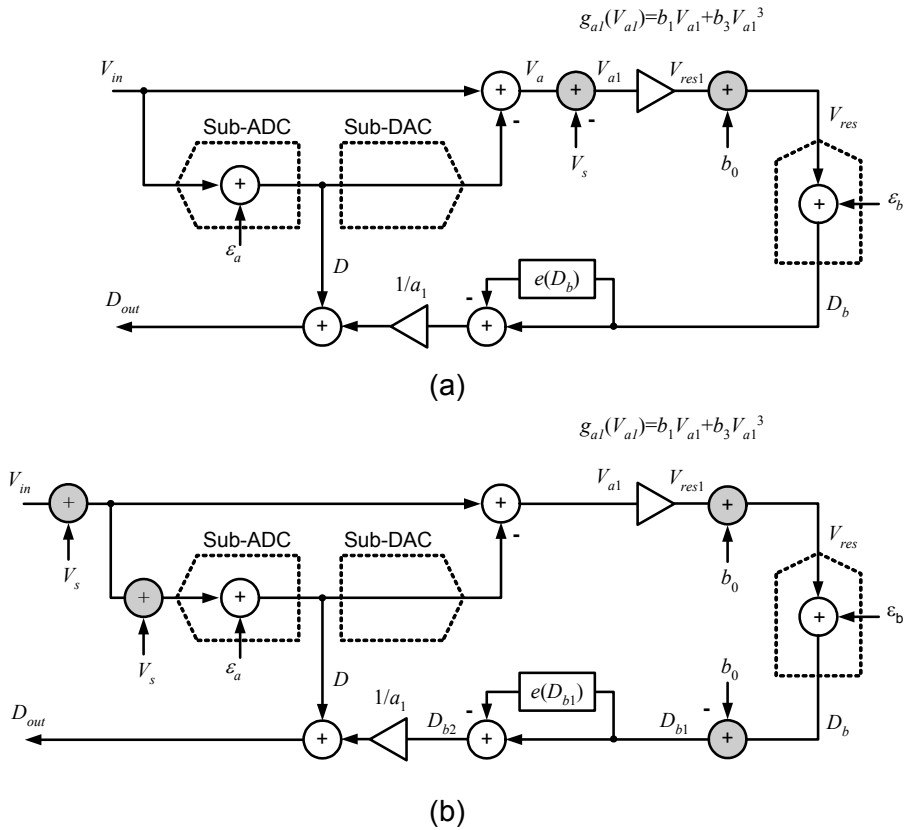
Figure 6-5. (a) Model with shifted variables. (b) Equivalent/compensated model.

The offsets V_s and b_0 are shown as shaded blocks and appear at the input and output of the residue amplifier respectively. As illustrated in Figure 6-5(b), the output referred offset is easily removed through a simple constant subtraction from the digital backend conversion result. The offset V_s can be pushed toward the input of the system, where it appears as an additional error in the sub-ADC and global input offset error. As we have seen above, sub-ADC errors do not affect the calibrated system as long as sufficient redundancy is present.

Two cases must be considered for the offset at the stage input. First, if the calibrated stage is the first stage of the pipeline, the offset represents a global ADC offset error, which is tolerated in most applications. Secondly, if the stage is located at some arbitrary location $j \in 2 \ldots n$ of the pipeline (see Figure 6-1(a)), V_s is indistinguishable from the output referred offset b_0 of stage j-1. In this case, the error will be absorbed by the calibration hardware

of stage j-1. Now, with respect to the new variable system, using $D_{b1}=D_b-b_0$, equation (6-9) becomes

$$e(D_{b1}) \cong D_{b1} - b_1 g_{a1}^{-1}(D_{b1}) .\qquad(6\text{-}13)$$

The inverse $g_{a1}^{-1}(D_{b1})$ of (6-12) can be found using trigonometric substitutions (see e.g. [100]). For the particular case of a gain compressive transfer function ($b_3/b_1 < 0$), the inversion and subsequent substitution into (6-13) yields

$$e(D_{b1}) \quad D_{b1} - 2\sqrt{\frac{-b_1^3}{3b_3}} \cos\left[\frac{\pi}{3} + \frac{1}{3}\cos^{-1}\left(\frac{D_{b1}}{2\cdot\sqrt{\frac{-b_1^3}{27b_3}}}\right)\right].\qquad(6\text{-}14)$$

A similar relationship can be found for the case of gain expansion ($b_3/b_1 > 0$), which is less common in electronic circuits. Despite its seemingly complex form, equation (6-14) can be implemented efficiently in hardware. The additive correction value $e(D_{b1})$ depends only on the backend data and the ratio b_3/b_1^3 as a single parameter. As discussed in chapter 8, a simple pre-computed 2-dimensional look-up table approach requires only a fairly small amount of memory.

As a final step in establishing the proposed correction block, it should be noted that the linear digital weighting by $1/a_1$ (Figure 6-5) is most efficiently achieved by scaling the sub-ADC result, rather than the digital backend code. This transformation is illustrated in Figure 6-6(a)-(c). In the final block diagram of Figure 6-6(c), the backend data is scaled by a power of two (2^R), which simply corresponds to a bit-shift of the binary backend data. The deviation of a_1 from the desired and ideal stage gain of 2^R is compensated by subtracting a small fraction of the sub-conversion result D. Since the bit-width of D is usually small, the required hardware overhead for this operation is low.

As a last simplifying step, we discard the final scaling of $2^R/a_1$. Just as in the case of the input referred stage offset discussed above, two cases must be considered to validate this step. First, if the stage under consideration is the first stage of the pipeline, the missing scaling operation represents a global ADC gain error that is tolerable in most applications.
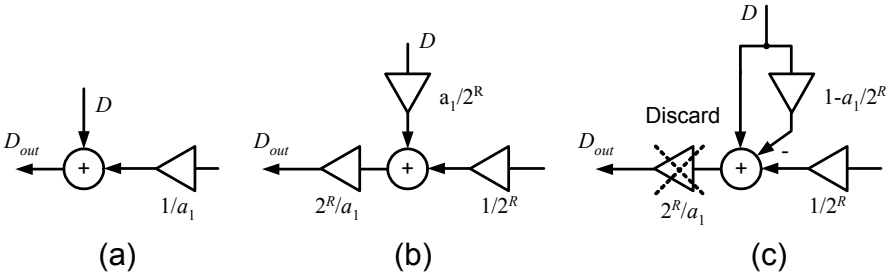
Figure 6-6. Modification for hardware efficient linear digital weighting.

Secondly, if the stage is located at some arbitrary location $j \in 2 \ldots n$ of the pipeline (see Figure 6-1(a)), the gain error can be lumped with the amplifier model of stage j-1 and will be compensated by the calibration hardware at this location.

Figure 6-7 summarizes the complete model for the digital correction blocks used in Figure 6.1.
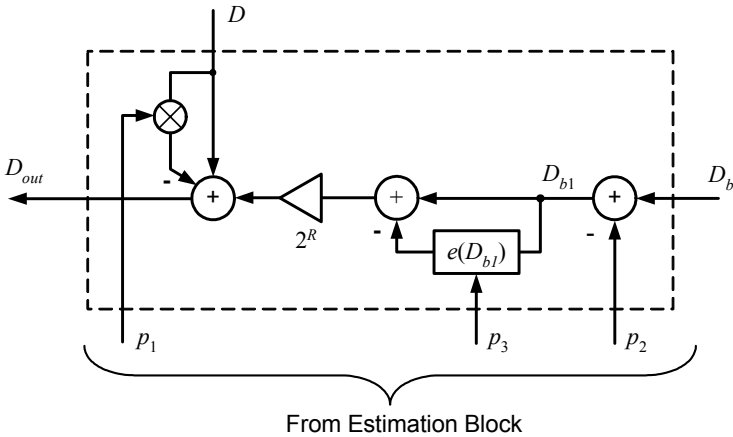
Figure 6-7. Complete digital correction hardware.

As derived above, the correction for linear, quadratic and cubic errors is based on a total of three parameters, shown as $p_1 \ldots p_3$ in Figure 6-7. The optimum values for these parameters are given by

$$p_{1opt} \quad 1 - \frac{b_1}{2^R} \quad 1 - \frac{a_1 - \frac{a_2^2}{3a_3}}{2^R}$$

$$p_{2opt} \quad \frac{2a_2^3}{27a_3^2} - \frac{a_1 a_2}{3a_3} \cong -\frac{a_1 a_2}{3a_3} \quad .$$

$$p_{3opt} \quad \frac{b_3}{b_1^3} \quad \frac{a_3}{\left(a_1 - \frac{a_2^2}{3a_3}\right)^3} \cong \frac{a_3}{a_1^3}$$

$$(6\text{-}15)$$

In practice, the amplifier coefficients $a_1...a_3$ in the above expressions are not precisely known and may also drift substantially over time and varying operating conditions. The digital background calibration algorithm described in the next chapter was designed to precisely estimate, and continuously update $p_1...p_3$ without interrupting normal converter operation.

3. ALTERNATIVE ERROR MODELS

As we have seen, a restriction to third only nonlinearity compensation results in a compact, low complexity digital correction scheme. In principle, higher order nonlinearities could also be compensated, with the potential advantage of mitigating the linearity-power tradeoff discussed in chapter 5.

For higher order compensation, however, the above algebraic approach seems unfeasible and complex. For a scheme that involves the compensation of higher order errors, it may be advantageous to consider nonlinearity models based on orthogonal representations, such as Chebychev polynomials (see e.g. [101]).

Chapter 7

STATISTICS-BASED PARAMETER ESTIMATION

1. INTRODUCTION

In the following sections, we describe a calibration technique that can be used to measure and track the digital correction parameters introduced in the previous chapter. In this technique, digital pseudo random modulation is used to identify and track amplifier nonlinearities in the "background," allowing the system to track device and environmental variations without interrupting normal ADC operation.

Background calibration of monolithic ADCs has been a popular research topic since the mid-1990s [102]. In previous work, it is often argued that the key advantage of a continuous calibration mechanism is its transparency to the user, who no longer needs to schedule calibration cycles that would interrupt normal ADC operation. In the proposed open-loop converter, the calibration coefficients relate to temperature sensitive amplifier coefficients that may drift substantially in short time intervals, which strictly dictates the implementation of a continuously tracking compensation approach.

A particularly interesting property of the technique described here is that it does not require generating an analog domain test signal, unlike other background calibration approaches. Instead, the calibration uses signal statistics, in a manner similar to the technique described in [26]. Conceptually, the estimation uses the fundamental property that perfectly linear systems at most scale, but never distort a signal's amplitude distribution. Deviations from this ideal case can be used to obtain information about the presence and magnitude of any nonlinearity. In some

sense, the (arbitrary) input amplitude distribution of the converter assumes the role of the calibration test signal.

2.   MODULATION APPROACH

The proposed statistics-based parameter estimation makes use of the fact that errors made in sub-ADCs do not affect the conversion result of a perfectly calibrated pipeline (equations (6-2) and (6-5)). This property invites a solution in which the response to a sub-ADC error modulation is used to estimate and minimize non-idealities.

For simplicity, consider first the system model with only linear gain correction as shown in Figure 7-1. Added to this model is an additive modulation to the digital output of the sub-ADC. If we let $MOD=+s/2$ and $MOD=-s/2$ respectively, we obtain for the digital output D_{out}

$$
\begin{aligned}
D_{out}^{(+s/2)} \quad & V_{in} + \left(\varepsilon_a + \frac{s}{2}\right)\cdot\left(1 - p_1 - \frac{a_1}{2^R}\right) + \frac{\varepsilon_b^{(+s/2)}}{2^R} \\
& V_{in} + \left(\varepsilon_a + \frac{s}{2}\right)\cdot\left(p_{1opt} - p_1\right) + \frac{\varepsilon_b^{(+s/2)}}{2^R}
\end{aligned}
\tag{7-1}
$$

$$
\begin{aligned}
D_{out}^{(-s/2)} \quad & V_{in} + \left(\varepsilon_a - \frac{s}{2}\right)\cdot\left(1 - p_1 - \frac{a_1}{2^R}\right) + \frac{\varepsilon_b^{(-s/2)}}{2^R} \\
& V_{in} + \left(\varepsilon_a - \frac{s}{2}\right)\cdot\left(p_{1opt} - p_1\right) + \frac{\varepsilon_b^{(-s/2)}}{2^R}
\end{aligned}
\quad .
\tag{7-2}
$$

Assuming that the input voltage V_{in} and thus also ε_a are constant in both modulation states, subtracting (7-2) from (7-1) yields

$$
\Delta D_{out} \quad D_{out}^{(+s/2)} - D_{out}^{(-s/2)} \quad s\cdot\left(p_{1opt} - p_1\right) + \frac{\varepsilon_b^{(+s/2)} - \varepsilon_b^{(-s/2)}}{2^R} .
\tag{7-3}
$$

Equation (7-3) indicates that this differential measurement is minimized for $p_1 = p_{1opt}$. In principle, and without prior knowledge of s and p_{1opt}, one could implement a deterministic, gradient-based algorithm that minimizes (7-1) over a sequence of measurements with constant V_{in}.
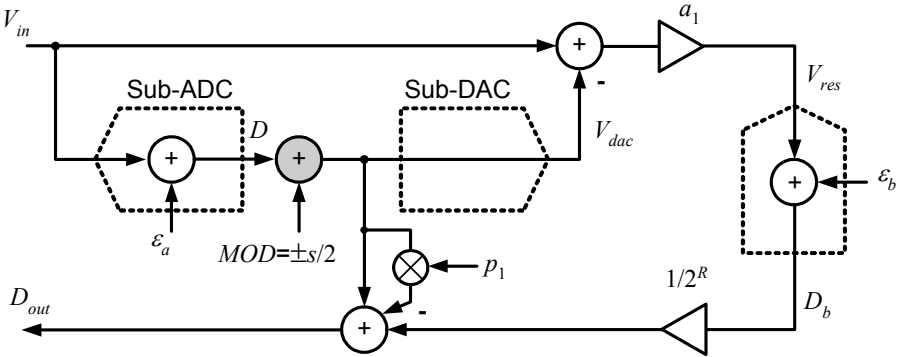
Figure 7-1. System model with digital code modulation.

Alternatively, to enable continuous background calibration of p_1 under varying V_{in}, we can choose the modulation signal such that

$$MOD(k) \quad (-1)^{RNG(k)} \cdot \frac{s}{2}, \tag{7-4}$$

where k is the discrete time index of the system, and $RNG(k) \in \{0,1\}$ is a binary random number generator sequence that is assumed to be uncorrelated with V_{in}. Under this condition, it is straightforward to show that the resulting output sequence $D_{out}(k)$ contains a term that is correlated with $MOD \cdot (p_{1opt} - p_1)$ and hence provides the desired optimization gradient in a statistical manner [28].

In the remainder of this chapter, we detail an extension to the above modulation principle that allows continuous background estimation of the all three parameters $p_1...p_3$.

3. REQUIRED SUB-ADC AND SUB-DAC REDUNDANCY

From the setup of Figure 7-1, we see that with the modulation applied, the peak input signal to the backend portion of the converter becomes

$$\left| V_{res} \right|_{\max} \quad a_1 \left(\left| \varepsilon_a \right|_{\max} + \frac{s}{2} \right) \tag{7-5}$$

From this expression, it is clear that without any further modifications, the modulation will require an excess dynamic range of $s/2$ in the backend, which may result in a significant power penalty for large s.

An elegant and efficient way to overcome this problem is to reduce the maximum quantization error $|\varepsilon_a|_{max}$ such that the overall peak magnitude of V_{res} does not increase compared to an unmodulated system. For instance, this can be accomplished by increasing the sub-ADC resolution by one bit, which amounts to a negligibly small power and area overhead in a practical pipeline design [85].

The proposed approach is illustrated in Figure 7-2 for the example of a 2-bit sub-ADC. Adding an extra bit reduces the quantization error by a factor of two. The extra headroom that is now available for modulation corresponds to $\pm\frac{1}{2}$ LSB of the 3-bit quantizer ($\pm\Delta/2$ in Figure 7-2(b)). Using this entire range for modulation by choosing $s=\Delta$, translates into a simple hardware implementation and also proves to be imperative for maximizing the signal-to-noise ratio of the estimation.

Figure 7-3(a) shows the resulting sub-ADC/DAC interface. As illustrated in the equivalent model of Figure 7-3(b), the introduction of $\frac{1}{2}$ LSB offset in the sub-DAC reduces the modulation to a simple conditional addition of 1 LSB (Δ) or a digital "1" to the sub-ADC output code. Assuming that the sub-ADC has a resolution of B_a bits and hence 2^{B_a} distinct output codes, the sub-DAC needs to provide $2^{B_a}+1$ output levels, due to the random addition of one LSB. This corresponds to slightly more than twice the number of levels of a conventional implementation without digital modulation. We shall see in chapters 8 and 9 that this modification represents only minor overhead and does not sacrifice performance. The required DAC unit element precision remains essentially the same, since tolerable errors are dictated by the backend resolution, rather than the local DAC resolution.
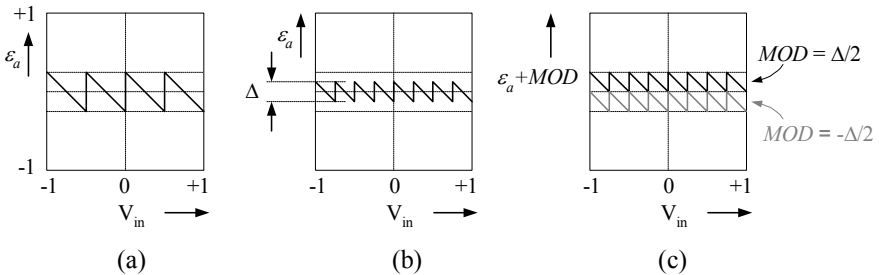

Figure 7-2. Introducing Sub-ADC redundancy: (a) Quantization error of a 2-bit sub-ADC. (b) Error of a (2+1)-bit sub-ADC. (c) Superimposed modulation.
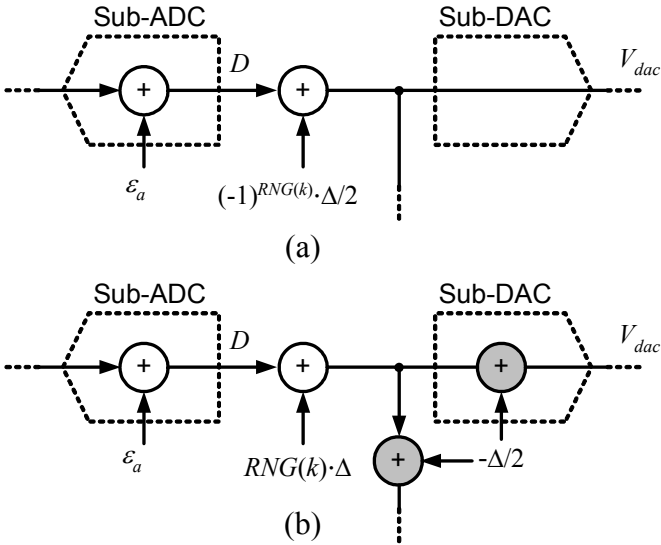
Figure 7-3. Sub-ADC/DAC interface: (a) Bipolar modulation.(b) Equivalent unipolar modulation with DAC offset.

It should be noted that the redundancy introduced for modulation does not help accommodate threshold errors in the sub-ADC. In a practical design, additional redundancy must be provided through one of the three approaches mentioned in section 1.2 of chapter 6.

4.   PARAMETER ESTIMATION BASED ON RESIDUE DIFFERENCES

In this section we construct a procedure to estimate parameters from function differences in the two modulation states. We examine the converter with respect to the commonly used stage residue plots, i.e. the transfer function from V_{in} to the output V_{res}, and its corrected digital representations D_{b1} and D_{b2} (Fig 6.5(b)). Figure 7-4 summarizes an appropriate model with the *RNG* modulation included for further discussion.

By the same reasoning as in chapter 6, we ignore the backend quantization error ε_b to simplify the analysis. For the two distinct states of the *RNG* signal, we obtain
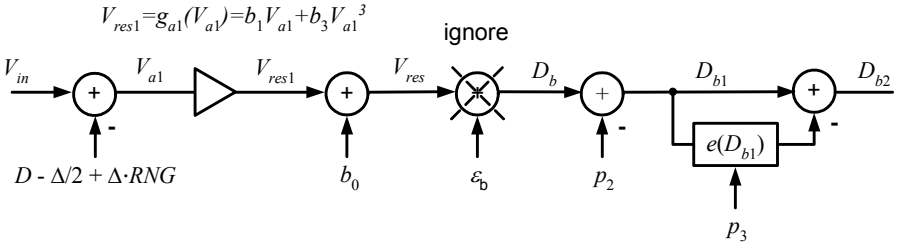
Figure 7-4. System model for transfer function analysis.

$D_{b2,0}$ $h_0(V_{in})$

$$g_{a1}\left(V_{in} - D + \frac{\Delta}{2}\right) + (b_0 - p_2) - e\left(g_{a1}\left(V_{in} - D + \frac{\Delta}{2}\right) + b_0 - p_2\right)$$

$D_{b2,1}$ $h_1(V_{in})$

(7-6)

$$g_{a1}\left(V_{in} - D - \frac{\Delta}{2}\right) + (b_0 - p_2) - e\left(g_{a1}\left(V_{in} - D - \frac{\Delta}{2}\right) + b_0 - p_2\right)$$

Since $V_{in}-D=-\varepsilon_a$, the argument of both functions is the (negative-) saw tooth quantization error function of the quantizer (Figure 7-2(b)), which is periodic with period Δ. Also note that

$$h_1(V_{in}) \quad h_0(V_{in} - \Delta) \tag{7-7}$$

The resulting two transfer functions are illustrated in Figure 7-5 for $B_a=2$ and the simple case of a perfectly linear system ($b_0=b_3=0$) with no correction applied ($p_2=p_3=0$). Each segment of this characteristic corresponds to a discrete value of the sub-ADC output D.

In the notation of this chapter, the discrete levels of D are given by

$$D \quad -1+\Delta\left(\frac{1}{2}+j\right); \quad j \quad 0,1,...2^{B_a} - 1$$

(7-8)

$$\Delta \quad \frac{2}{2^{B_a}}$$

Without loss of generality, we now focus on one segment of this characteristic, and choose for notational convenience $j \quad 2^{B_a}/2$, which corresponds to $D = \Delta/2$. Figure 7-6 shows this transfer function segment in the presence of nonlinearities.
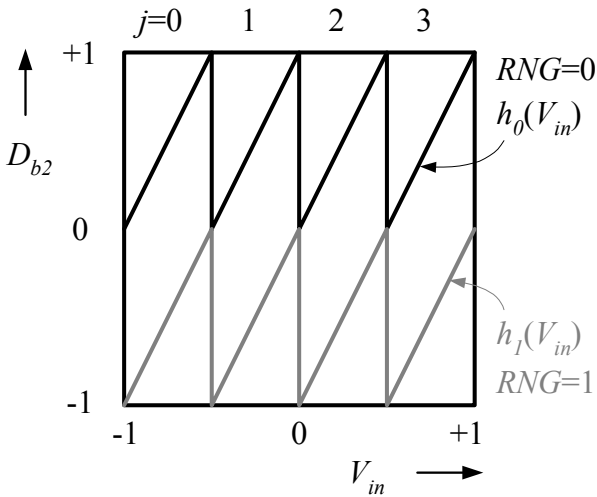
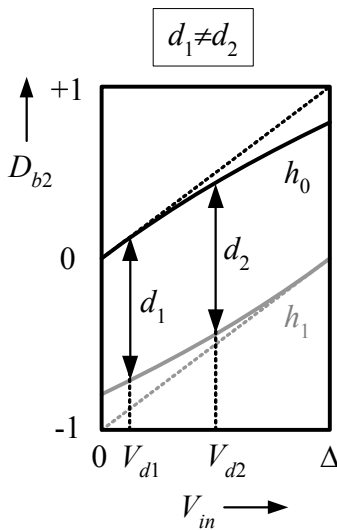*Figure 7-5.* Residue plot for both *RNG* states.

Figure 7-6. Single transfer function segment without correction and $b_3 < 0$, $b_0 = 0$.

First, consider the case with only cubic amplifier distortion and no correction applied. Equation (7-6) then simplifies to

$$h_0(V_{in}) \quad b_1 V_{in} + b_3 V_{in}^3$$
$$h_1(V_{in}) \quad b_1(V_{in} - \Delta) + b_3(V_{in} - \Delta)^3 \tag{7-9}$$

Annotated in Figure 7-6 are the residue differences d_1 and d_2 for two fixed input voltages (V_{d1} and V_{d2}) near the center and edge of the segment respectively. Mathematically, the difference between these two quantities is

$$\Delta d \quad d_1 - d_2 \quad 3b_3 \Delta \left[(V_{d1} - V_{d2})(\Delta - (V_{d1} + V_{d2})) \right]. \tag{7-10}$$

As we see from this expression and also graphically from Figure 7-6, the difference between the two measurements vanishes for $b_3=0$, i.e. for a perfectly linear amplifier. Alternatively, this could also be achieved for $b_3 \neq 0$ but with active, perfectly adjusted digital correction that maps both residues onto straight lines. If the correction function $e(D_{b1})$ is applied, we find

$$\Delta d \quad d_1 - d_2 \cong 3b_1^3 \Delta \cdot (p_3 - p_{3opt}) \cdot \left[(V_{d1} - V_{d2})(\Delta - (V_{d1} + V_{d2})) \right]. \tag{7-11}$$

This result indicates that the deviation of parameter p_3 from its ideal value is directly proportional to Δd. In principle, this gradient information could be used in a search algorithm that minimizes (7-11) and thus optimizes p_3 over a sequence of measurements with constant test voltages V_{d1} and V_{d2} applied in both modulation states. Section 5 introduces a statistics-based difference estimation approach that avoids the need for constant inputs and therefore allows calibration in the background, during normal converter operation.

To refine the idea of parameter calibration based on residue distances, consider now choosing the measurement locations of a second set of differences based on symmetric ordinates (see e_1 and e_2 in Figure 7-7). The ordinates y_{e1} and y_{e2} are chosen such that $y_{e2}=-y_{e1}$, so that $V_{e1}=h_0^{-1}(y_{e1})$ and $V_{e2}=h_1^{-1}(-y_{e1})$. Using (7-6) and the general transfer function of the model in Figure 7-4, we find

$$\Delta e \quad e_1 - e_2 \quad h_0\left(h_0^{-1}(-y_{e1}) + \Delta\right) + h_0\left(h_0^{-1}(y_{e1}) - \Delta\right). \tag{7-12}$$

This expression equals zero if and only if h_0, and consequently also its inverse h_0^{-1} are odd functions ($h_0(V_{in})=-h_0(-V_{in})$, $h_0^{-1}(y)=-h_0^{-1}(-y)$). Since h_0 is odd if and only if the quadratic error term is perfectly cancelled, a vanishing Δe indicates perfect calibration ($p_2=p_2opt$).
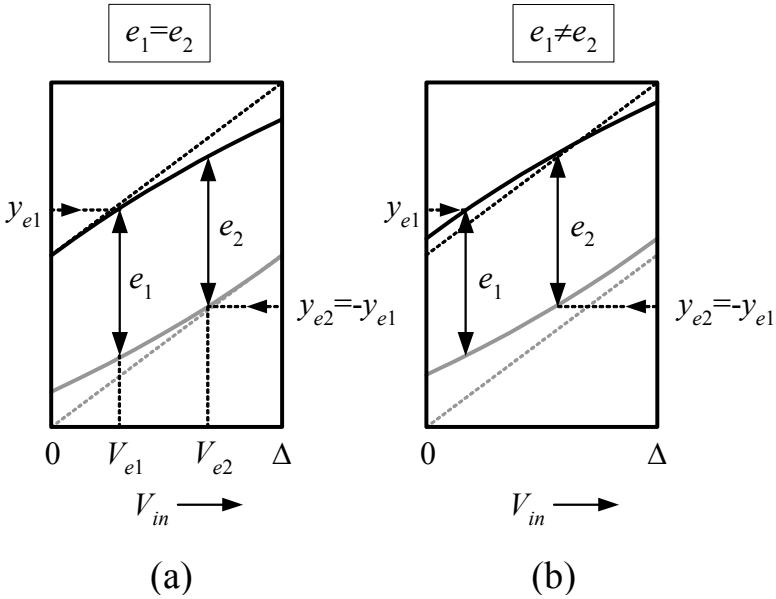
Figure 7-7. Difference measurement with symmetrical ordinates ($b_3 < 0$, $b_0 = 0$). (a) Symmetry with ($b_0 = 0$). (b) Asymmetry caused by $b_0 \neq 0$.

This is also seen graphically in Figure 7-7. With only cubic distortion present (Figure 7-7(a)), point symmetry around ($\Delta/2$, 0) results in $e_1 = e_2$, independent of the amount of cubic distortion[2]. With a quadratic component ($a_2 \neq 0 \Rightarrow b_0 \neq 0$) the point symmetry is lost and results in a difference between the two measurements. Using suitable approximations, and assuming weak nonlinearity, we find

$$\Delta e \quad e_1 - e_2 \cong -3\frac{b_3}{b_1}\Delta\left(\Delta - \frac{y_{e1}}{4}\right)(p_2 - p_{2opt}). \qquad (7\text{-}13)$$

Hence, the two distance measurements based on symmetric ordinates provide a suitable gradient for calibrating the parameter p_2. Once p_2 and p_3 are perfectly adjusted, all residue curves are mapped onto perfectly straight lines with slope b_1, and therefore

$$d_1 \quad d_2 \quad e_1 \quad e_2 \quad b_1\Delta, \qquad (7\text{-}14)$$

---

[2] Mathematically, this is confirmed by the trivial root in (7-11), i.e. $\Delta d = 0$ for $(V_{d1} + V_{d2})/2 = \Delta/2$, independent of b_3.

independent of measurement location. Assuming a perfect sub-DAC, we have $\Delta \approx 2/2^{B_a}$, and thus e.g.

$$p_{1opt} \approx 1 - \frac{b_1}{2^R} \approx 1 - \frac{d_1}{2} \cdot \frac{2^{Ba}}{2^R}. \tag{7-15}$$

Therefore, the optimal value of the calibration parameter p_1 can be directly obtained from one or several distance measurements at any location. In the proposed implementation (see chapter 8), the difference estimates for cubic calibration are re-used to obtain p_1. Alternatively, one could estimate this parameter with separate hardware, based on the correlation principle used in [28].

5.   STATISTICS BASED DIFFERENCE ESTIMATION

Figure 7-8 illustrates the proposed statistics-based residue difference measurement, which does not require constant or known inputs. In the following discussion we focus on the estimation of a single residue difference in one transfer function segment. As a further simplification, we assume that $V_{in}(k)$ is a stationary, "white", discrete time random process, whose samples are described by a well behaved, but otherwise arbitrary probability density function (PDF) .

The proposed distance estimation is based on evaluating cumulative histograms of the digital backend data (D_{b1} or D_{b2}). Figure 7-8(a) reviews the basic concept of a cumulative histogram. In this simple example, we consider only the bottom residue curve (fixed $RNG=1$) and one histogram bin at a particular code location y_{bot}.

The cumulative histogram count $M=CH(y_{bot})$ is found by counting the number of samples seen in the backend that are less than or equal to the reference code y_{bot}. Hence, the expected value of $CH(y_{bot})$ will be proportional to the total number of samples processed times the hatched area underneath the PDF, which represents the probability of an input sample being below the code threshold V.

With the RNG switching randomly, one of the two residue curves is chosen for each sample with equal probability and independent of V_{in}. Consider now a second cumulative code bin $CH(y_{top})$ that is associated with the top residue, as shown in Figure 7-8(b). For the time being, assume that the decision level of code y_{top} precisely coincides with V (the decision level of code y_{bot}).
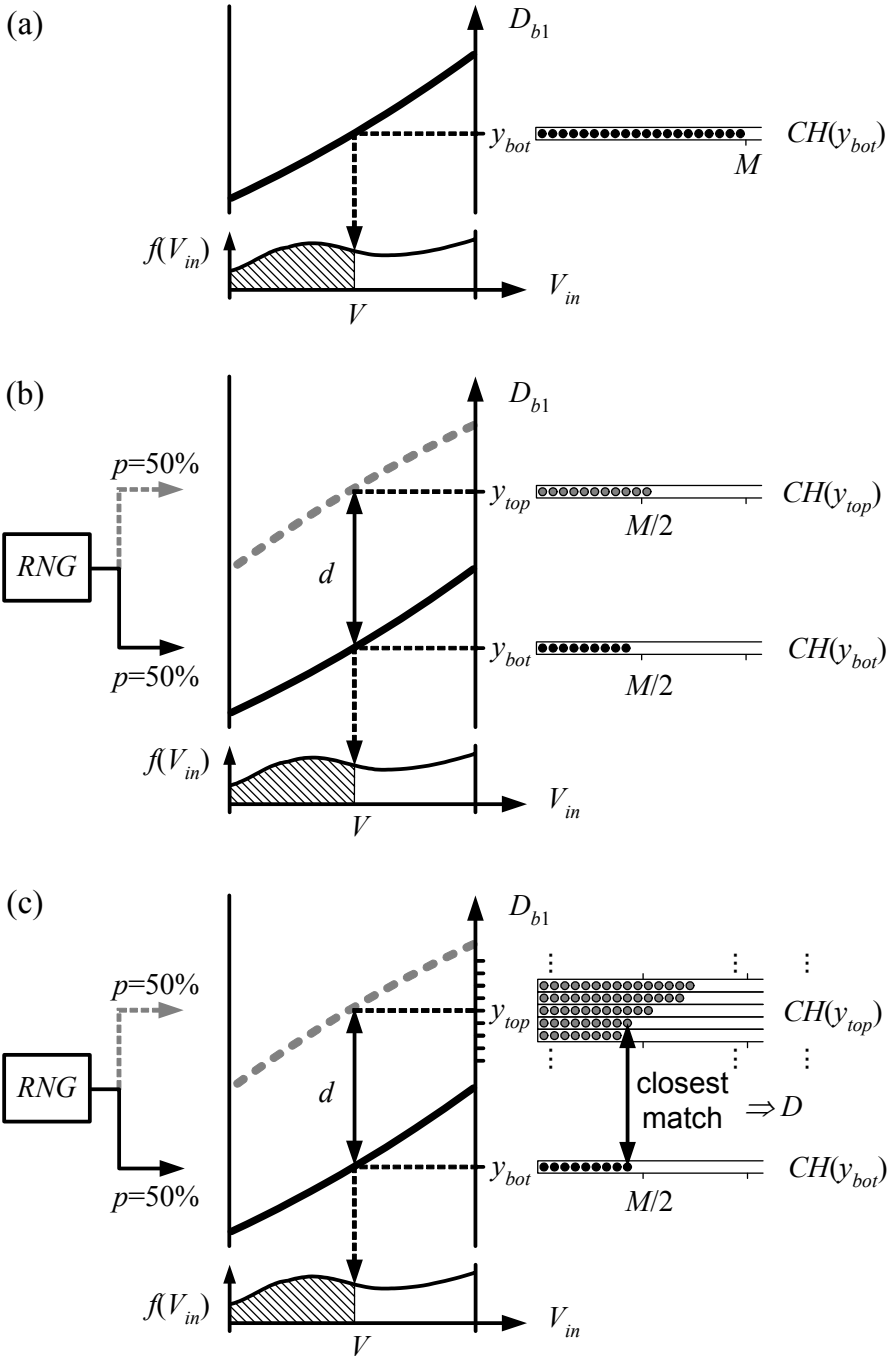
Figure 7-8. Statistics based distance estimation. (a) Cumulative count with *RNG* fixed. (b) Random split with active *RNG*. (c) Distance estimate from closest cumulative count.

Due to the modulation, the count $CH(y_{bot})$ of Figure 7-8(a) is now split into two histogram bins. Detailed analysis shows that the expected value in each bin is $M/2$, but due to randomness in the modulation, particular outcomes vary and typically won't result in a perfect $M/2$ split. This fact is illustrated as slightly imbalanced counts in Figure 7-8(b).

Consider now the setup of Figure 7-8(c), in which several additional cumulative code bins have been added around code y_{top}. With the random modulation in progress, and after processing a large number of samples N, the top bins are evaluated and compared to the bottom count $CH(y_{bot})$. From the closest match, it is straightforward to obtain the distance estimate D (see also definitions (B-1) and (B-4) in appendix B). It can be shown that the random variable D is an asymptotically unbiased estimate of the true residue distance d, i.e. for increasingly large N, the expected value of the estimate approaches the true value. The detailed analysis in appendix B shows that

$$\mathrm{var}(D) \cong \frac{4}{N} \cdot \frac{F(V)}{f(V)^2} , \qquad\qquad (7\text{-}16)$$

where $f(V)$ and $F(V)$ denote the probability density and cumulative distribution functions of the input samples $V_{in}(k)$, evaluated at the estimation site V. For the special case of a uniform input distribution, and letting $b_1 = 1/\Delta$ (full-swing residues with no redundancy as in Figure 7-5), (7-16) becomes

$$\mathrm{var}(D) \cong \frac{4}{N} \cdot \frac{V}{\Delta} . \qquad\qquad (7\text{-}17)$$

Qualitatively, and from the derivation in the appendix, it is clear that this equation does not hold for $V=0$, i.e. placement of an estimator at the segment edge. Moreover, this choice is impractical, since there is uncertainty in the segment boundaries due to sub-ADC noise and offset. From a practical perspective, there exists a reasonable, minimum choice for V that is commensurate with the expected sub-ADC precision of the implementation. In all further derivations, we refer to this quantity as V_{min}.

At first glance, equation (7-17) also seems impractical since in most applications the ADC input may not be uniformly distributed. Especially in communications systems, channel coding schemes tend to generate signals with approximately Gaussian distributions. However, if the histograms of Figure 7-8 are taken from the combined backend data[3], without

---

[3] Note that using combined data for the estimation requires that all segments be described by identical power series. See also discussion in appendix A.

distinguishing between segments, the effective distribution seen by the histogram bins is the average of the individual distribution segments. This averaging effect tends to produce net distributions that are fairly uniform. Figure 7-9 shows an example for a Gaussian input and four combined segments. The implementation considered in this book is based on binning samples from all segments using one single histogram. Only samples from the bottom and top transfer curves are processed separately as required by the algorithm. This justifies using (7-17) for all further considerations.

6.   COMPLETE ESTIMATION BLOCK

Combining the concepts above, we now construct a suitable realization of the complete estimation block used in Fig 6-1. Figure 7-10 shows a block diagram of the proposed system using adaptive least mean square (LMS) loops [103]. Here, the scheme described in the previous section is replicated to generate statistics-based estimates for the deterministic quantities d_1, d_2, e_1 and e_2. In Figure 7-10 these variables are denoted by their respective upper case symbol. In all three estimation loops, the presence of a discrete time integrator forces the mean value of the loop inputs to zero, which corresponds to optimum calibration. For the case of the linear calibration loop, the mean of the difference

$$\Delta p_1 \quad p_1 - \left(1 - \frac{D_1}{2} \frac{2^{Ra}}{2^R}\right) \quad\quad (7\text{-}18)$$
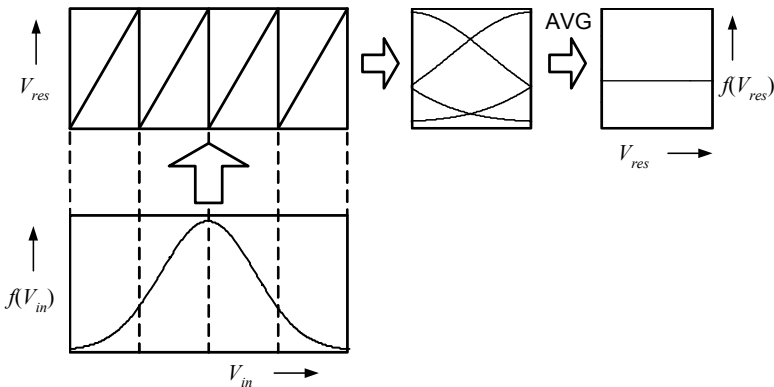
is forced to zero.
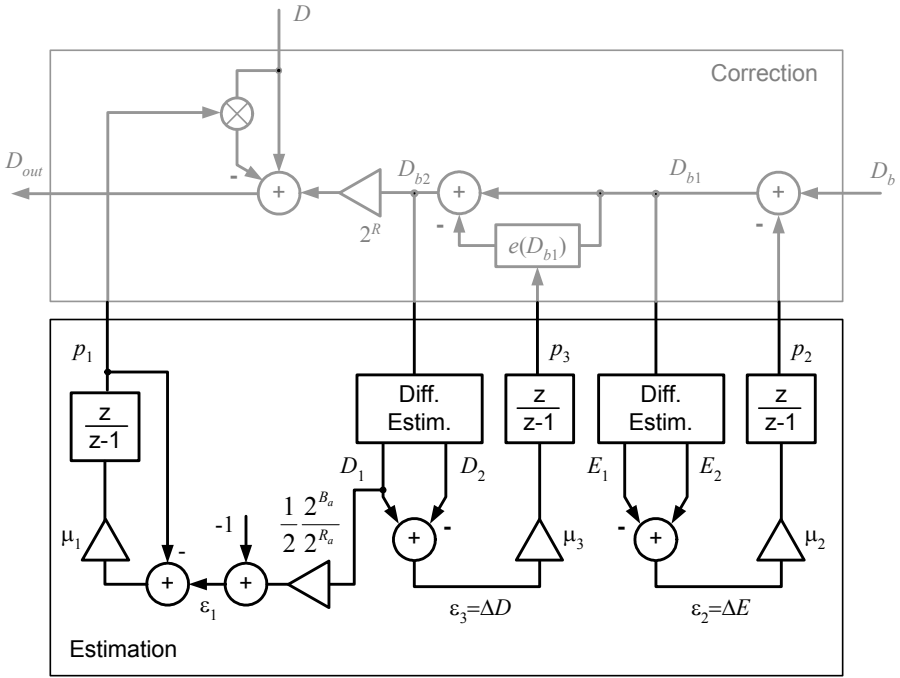

Figure 7-9. Averaging effect.

Figure 7-10. Parameter estimation using LMS loops.

We therefore have

$$E(p_1) \quad E\left(1 - \frac{D_1}{2}\frac{2^{Ra}}{2^R}\right) \quad 1 - \frac{d_1}{2}\frac{2^{Ra}}{2^R} \quad p_{1opt}. \tag{7-19}$$

Similarly, in both the quadratic and cubic calibration loops, the mean of the two difference estimates are forced to zero, which produces optimum estimates for p_2 and p_3 (see equations (7-11) and (7-13)).

Due to the statistical variations in the difference estimates, there exists a certain variance in the loop outputs $p_1 \ldots p_3$. The analysis in appendix C shows that

$$\text{var}(p_i) \cong \frac{1}{2}\frac{\mu_i}{\delta_i}\text{var}(\varepsilon_i), \tag{7-20}$$

where μ_i and ε_i are as indicated in Figure 7-10, and

$$\delta_i \cong \frac{d\varepsilon_i}{dp_i}. \tag{7-21}$$

By inspection of Figure 7-10, we see that $\delta_1=1$. For the quadratic and cubic loops, the δ terms can be found by differentiating (7-11) and (7-13).

From (7-20), it follows that the loop gain parameters μ_i should be chosen as small as possible to minimize inaccuracy in the correction parameters. For given precision requirements in each one of the three LMS loops, this translates into an upper bound for the loop gain parameters of the form

$$\mu_i \leq \frac{\delta_i}{\alpha_i} L_i^2 \left(\frac{2}{2^{B_b}} \right)^2 N, \tag{7-22}$$

where N is the number of samples processed until histogram evaluation, B_b is the effective resolution of the converter backend, and L_i quantifies the worst case DNL error budget allocated in each loop in LSBrms. The loop specific parameters α_i are given by

$$\begin{array}{cc} \alpha_1 & 2\dfrac{V_{\min}}{\Delta} \\[2ex] \alpha_2 & 4\dfrac{V_{\min}}{\Delta} \\[2ex] \alpha_3 & 1+\dfrac{2\cdot V_{\min}}{\Delta} \end{array} \tag{7-23}$$

These constants capture the variance of the distance estimates in each loop as a function of their location (see derivation in appendix C).

Unfortunately, reducing the parameters μ_i increases the LMS loop time constants and therefore impairs the tracking capability of the system. For equality in (7-22) this translates into minimum attainable time constants given by

$$\tau_{i\min} \cong \frac{N}{f_s} \frac{1}{\mu_{i\max}\delta_i} \quad \frac{1}{f_s} \frac{\alpha_i}{\delta_i^2} \frac{1}{L_i^2} \left(\frac{2^{B_b}}{2} \right)^2, \tag{7-24}$$

where f_s is the sampling frequency of the converter and all other parameters are as discussed above. Note this result is independent of N, the number of

samples in one estimation cycle. Heuristically, we can argue that N should be chosen such that the standard deviation of the distance estimates corresponds to at least several LSBs (bin widths) of the backend quantizer. Under this condition, the estimator error is dominated by its inherent statistical variance, rather than the quantization noise of the backend. In some sense, the estimator variance then acts as a "dither" signal that reduces the relative impact of the finite granularity of the histogram bins. For the estimate D_1, using (7-17), this consideration translates into

$$\text{var}(D_1) \cong \frac{4}{N} \frac{V_{\min}}{\Delta} > \left(m \cdot \frac{2}{2^{B_b}} \right)^2 \tag{7-25}$$

and thus

$$N < \frac{V_{\min}}{\Delta} \frac{2^{2B_b}}{m^2}, \tag{7-26}$$

where m quantifies the expected bin span in LSBrms.

7.   SIMULATION EXAMPLE

In this section we illustrate the capabilities of the proposed calibration technique through numerical examples and simulation. As a demonstration vehicle, we use a simulation model that closely resembles the pipelined ADC implementation of chapter 8. This converter consists of a 3-bit first stage and a backend that has effective resolution of 9 bits plus two redundant bits for calibration purposes. In this example, only the multi-bit first stage is calibrated for errors caused by nonlinear open-loop residue amplification and all other stages are assumed to be perfect. An appropriate model for the first stage amplifier can be derived from (5-2) and is given by

$$g_a(V_a) \quad g_m R \cdot \left[\left(\frac{V_a}{V_{ref}} \right) + \frac{1}{4} \frac{\Delta\beta}{\beta} \left(\frac{V_{ref}}{V_{OV}} \right) \left(\frac{V_a}{V_{ref}} \right)^2 - \frac{1}{8} \left(\frac{V_{ref}}{V_{OV}} \right)^2 \left(\frac{V_a}{V_{ref}} \right)^3 \right]. \tag{7-27}$$

Table 7.1 below summarizes the associated design parameters and the values assumed in this example.

Table 7-1. Open-loop amplifier parameters.

Parameter	Description	Value
V_{ref}	Converter reference voltage	1 [V]
g_mR	Linear amplifier gain term	8 - 5% = 7.6
V_{OV}	Differential pair gate overdrive	0.25 [V]
$\Delta\beta/\beta$	Transistor mismatch	5%

With the values given in Table 7.1, (7-27) becomes

$$g_a(V_a) \quad 7.6V_a + 0.38V_a^2 - 15.2V_a^3. \qquad (7\text{-}28)$$

Figure 7-11 shows the converter's DNL and INL without any digital correction applied ($p_1=p_2=p_3=0$). As we see from the DNL signature, the negative gain error of the amplifier results in a large number of missing codes. The large amount of positive and negative INL is caused by these missing codes and also by the quadratic and cubic error terms in (7-28).
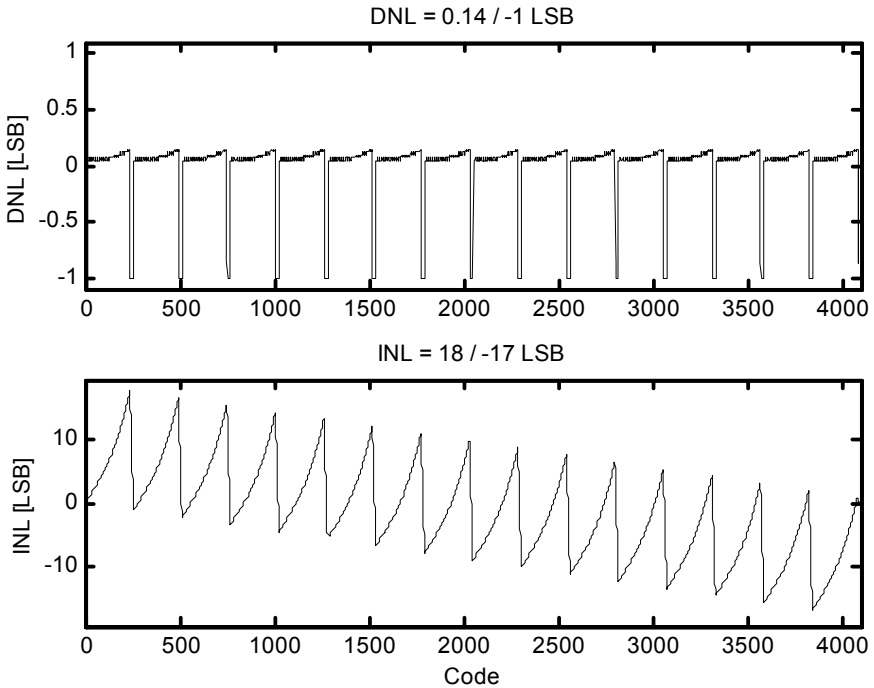

Figure 7-11. DNL and INL without correction (*RNG*=0).

With perfectly adjusted $p_1 \ldots p_3$ we obtain the nonlinearity plot shown in Figure 7-12. Both DNL and INL are now reduced to peak values of approximately 0.25 LSB and 0.3 LSB respectively. Two effects explain these small residual errors. First, re-quantization from the 14-bit raw data to the final 12-bit conversion result introduces an error. Secondly, the signal dependent slope change in the residue segments causes additional DNL. Both of these effects were discussed in more detail under section 2 of chapter 6.

This significant improvement in converter linearity can also be seen in the frequency domain. Figure 7-13 and 7-14 compare the results of a tone test with and without digital correction. With perfect calibration, the effective number of bits (ENOB) improves from 7.8 to 11.8, which is close to ideal converter operation.
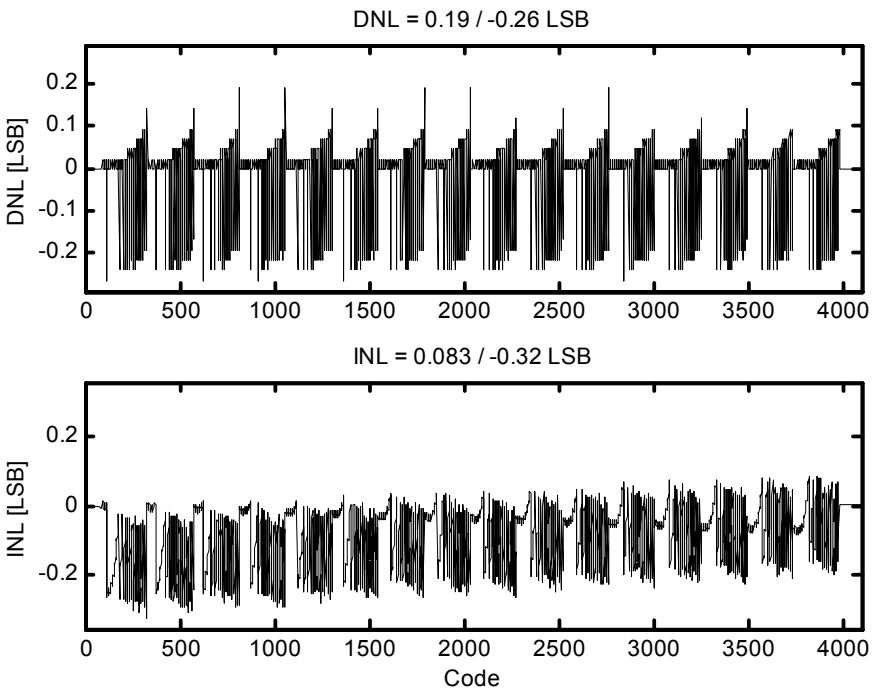

Figure 7-12. DNL and INL with perfectly adjusted calibration parameters (*RNG*=0).
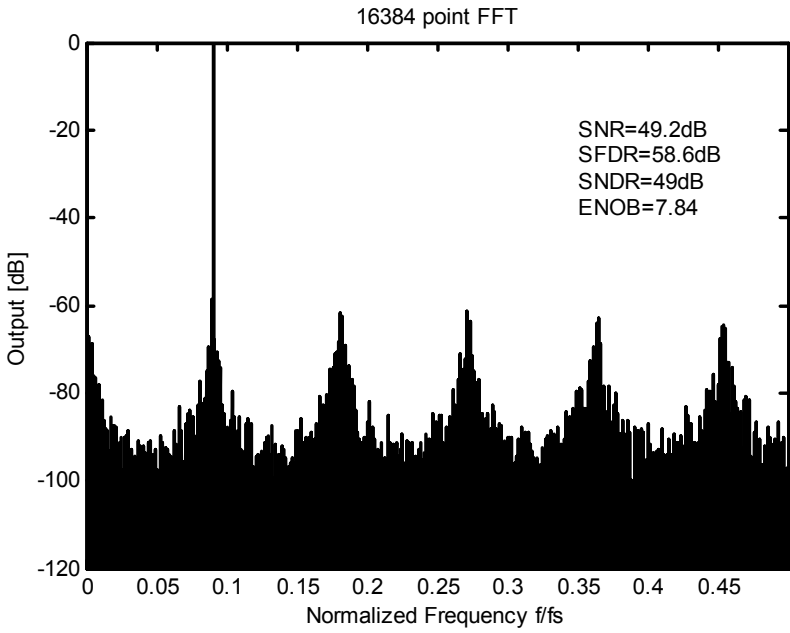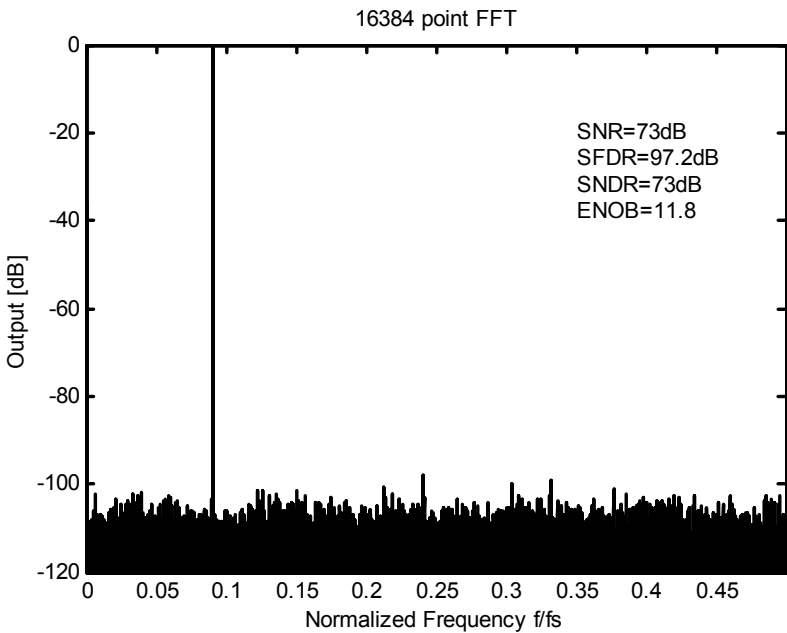
Figure 7-13. FFT without correction (*RNG*=0).

Figure 7-14. FFT with perfectly adjusted correction parameters (*RNG*=0).

Next, we include the LMS estimation loops in the simulation. Using $V_{min}=\Delta/8$, we obtain $\delta_2=0.136$ and $\delta_3=0.422$. From equation (7-24) we see that the expected loop time constant is inversely proportional to δ^2. In order to compensate for the low sensitivity δ_2, we allocate most of the total worst case DNL error budget to the second loop. With a total error budget of

$$L_{tot} \quad \sqrt{L_1^2 + L_2^2 + L_3^2} \quad 0.5 LSBrms, \tag{7-29}$$

and allocating approximately 80% of this budget for L_2^2, 15% for L_3^2 and the remaining 5% for L_1^2, we obtain the loop parameters summarized in Table 7-2 below.

Table 7-2. LMS Loop Parameters (*N*=30,000).

Gain Factors	Time Constants
$\mu_1=1/170$	$\tau_1=5.1\cdot10^6/f_s$
$\mu_2=1/40$	$\tau_2=8.8\cdot10^6/f_s$
$\mu_3=1/170$	$\tau_3=12\cdot10^6/f_s$

For these calculated values, we assumed a cycle length of *N*=30,000, i.e. 30,000 samples are collected until histogram evaluation.

Figure 7-15 shows the parameter convergence upon startup of the converter, with a full-scale sine wave applied. Parameter p_2 converges as expected from the time constant calculated above. The deviation of the other two parameters from their expected envelope is caused by the fact that the three estimation loops are not orthogonal. For instance, p_3 must first follow the transients in p_2 before it can reach its ideal steady-state value. Parameter p_1 must track the settling of both p_2 and p_3. Figure 3-13 shows the settling of p_1 with p_2 and p_3 in steady state (near optimum values), in which case the convergence occurs with the expected envelope.

Figure 7-17 below shows the effective number of bits (ENOB) during parameter settling. The ENOB reaches its steady state after roughly 40 Million samples. This number corresponds to about three time constants of the quadratic and cubic estimation loops (see Table 7-2). The distribution of the effective number of bits in steady state is shown in Figure 7-18. From the percentile plot, we see that the statistical nature of the estimation accounts for a worst case ENOB loss of about 0.15, compared to the maximum value. This penalty could be reduced to arbitrarily small values at the expense of larger LMS tracking time constants.
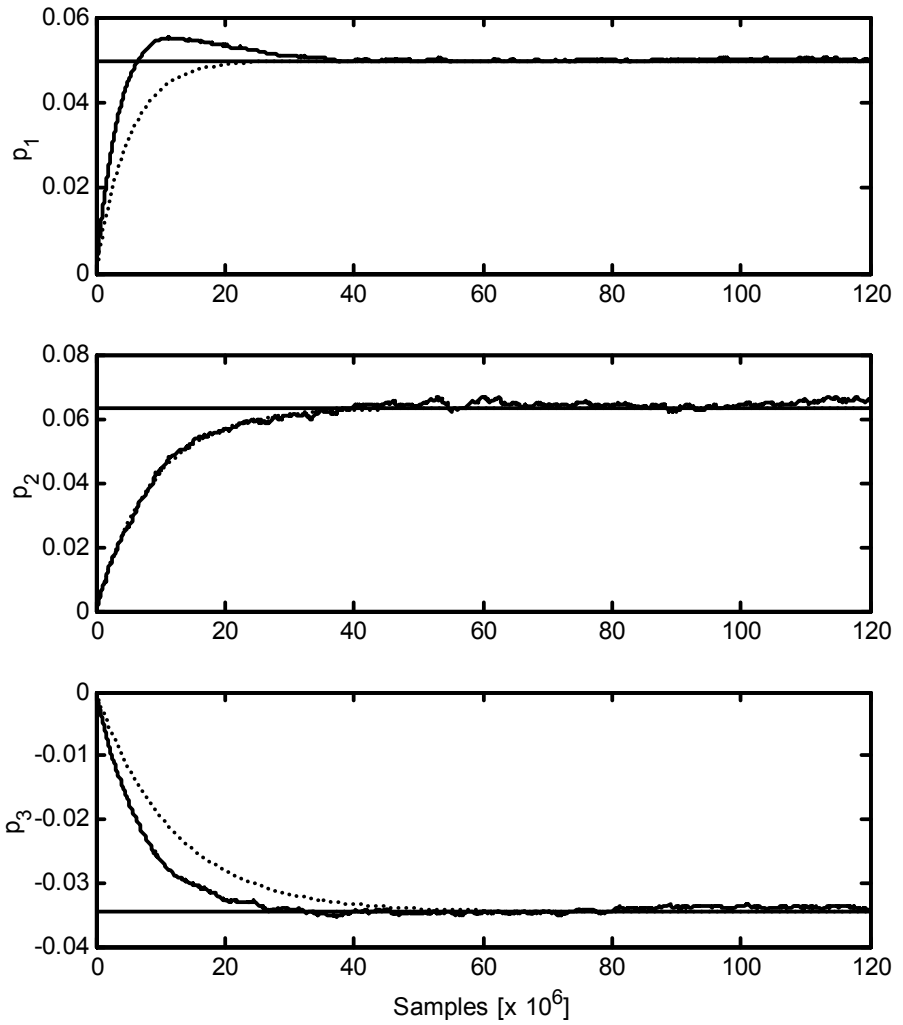
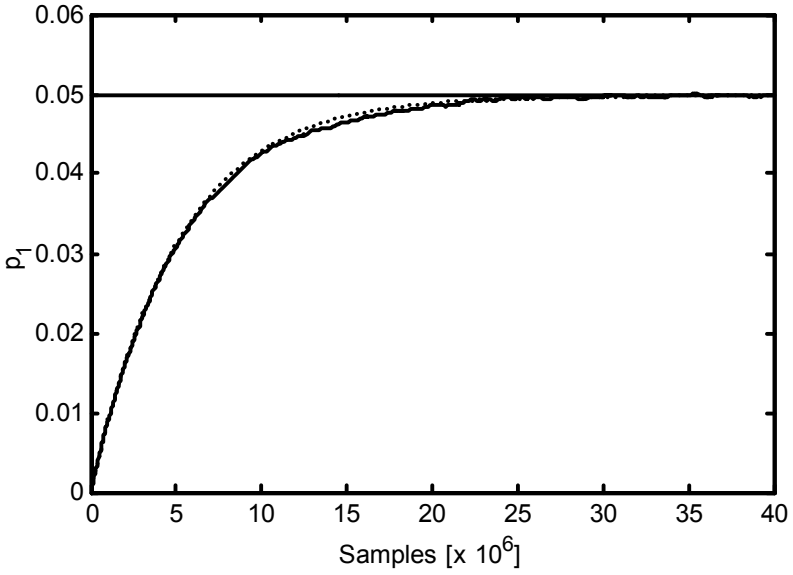*Figure 7-15.* Parameter convergence (dotted lines show the expected envelope).

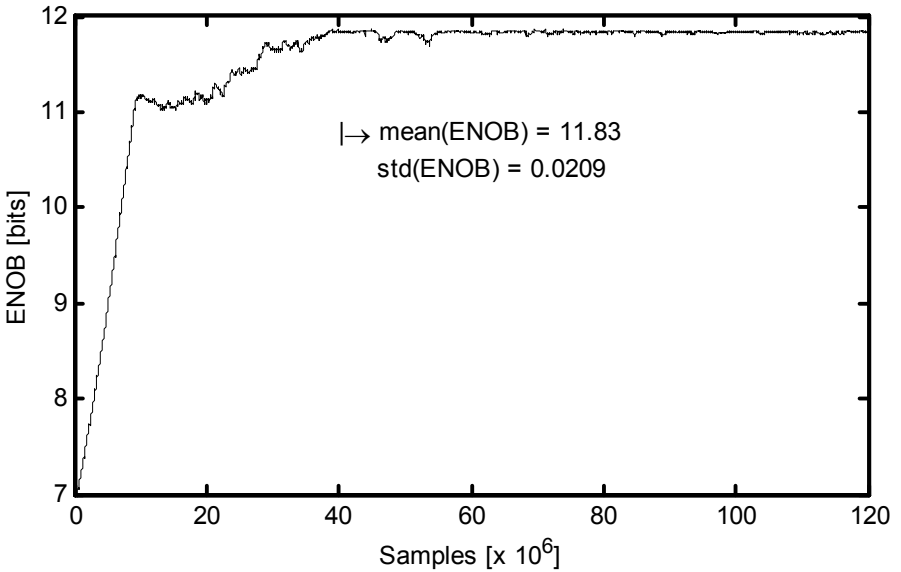*Figure 7-16. p_1 convergence with p_2 and p_3 in steady state.*
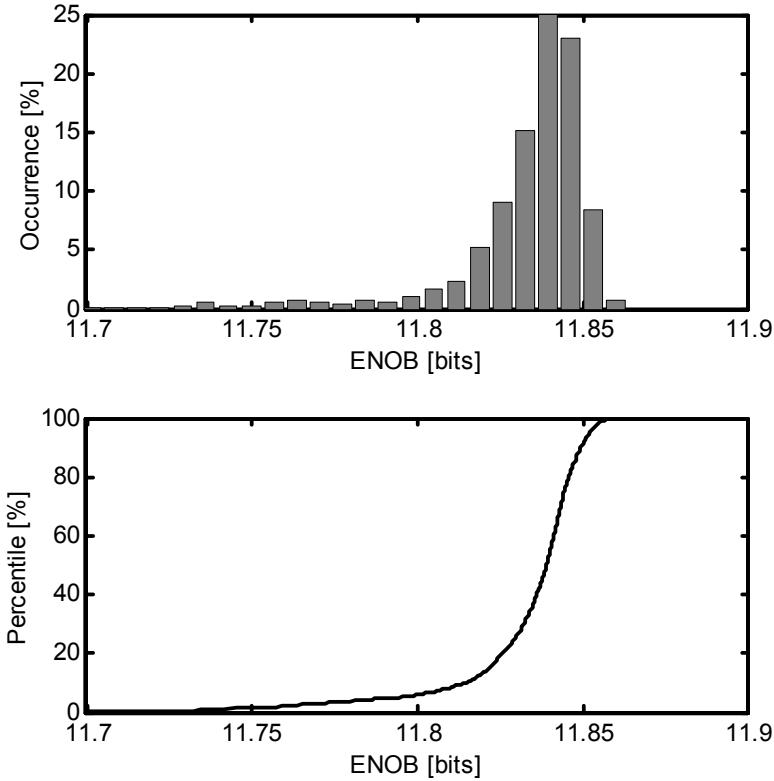

Figure 7-17. ENOB convergence.

Figure 7-18. ENOB distribution in steady state.

As a further consideration, the initial convergence time of the system could be reduced by dynamically varying the loop gain during start-up. For instance, the μ-parameters could be chosen large initially to achieve fast settling, and reduced later to reduce the steady-state variance in $p_1...p_3$. Such "variable step size" LMS algorithms have been studied extensively in literature [103].

8. DISCUSSION

8.1 Input Signal Limitations

From Figure 7-8, it is clear that the calibration algorithm fails if the input signal is not sufficiently "busy" around the input voltages at which the distance estimates are taken. Inactivity results in a flat cumulative histogram with indistinguishable bins in the top counter array. It can be argued that

this property only mildly affects the practicality of the proposed technique approach. First, insufficient amplitude activity can be easily detected, making it possible to avoid miscalibration due to low swing, quasi-DC input signals. Furthermore, since the estimation process combines backend data from several segments, activity spanning only a fraction of the converter's full-scale range is sufficient for calibration.

8.2   Tracking Time Limitations

As we have seen from the above implementation example, the statistical nature of the parameter estimation dictates fairly large tracking time constants. For the discussed 12-bit implementation, time constants on the order of 10 Million samples are necessary. Assuming a conversion rate of 100MS/s in a typical state-of-the-art ADC, this translates to 100ms on an absolute time scale.

For ADC resolutions of 8-12 bits, the attainable tracking speed is sufficient to compensate e.g. ambient temperature variations, slow changes in supply voltage and device aging effects. Potentially faster variations, for instance due to self-heating effects, must still be addressed by appropriate analog circuit design techniques. Measures to reduce the sensitivity of the open-loop ADC to potentially faster variations are briefly discussed in chapter 8.

For higher resolution ADCs, e.g. 14 or 16 bits, the required time constants become very large. From (7-24), we see that each additional bit in ADC precision results in quadrupling the time constant. Hence, for a 16-bit converter, we would expect time constants of 100ms$\cdot 4^4$ or roughly 26 seconds. In cases where such slow adaptation cannot be tolerated, a modified estimation process that uses a "split-ADC" approach could be considered [28].

8.3   DAC Error Compensation

In the proposed digital correction and parameter estimation, we assumed perfect sub-DAC operation. In an implementation where DAC errors are critical, they can also be corrected digitally, in a very similar way to the correction of linear gain errors through parameter p_1. Analysis shows that keeping separate linear correction parameters for each DAC state is sufficient for perfect error cancellation (see e.g. [99]).

In principle, one could augment the proposed scheme such that DAC correction parameters are also calibrated in the background. However, since DAC errors are usually given by component mismatch, which does not drift significantly over time, a simple one-time, "foreground" calibration should

suffice. Suitable start-up DAC calibration techniques for pipelined ADCs have been described for example in [99, 104, 105].

This page intentionally left blank

Chapter 8

PROTOTYPE IMPLEMENTATION

This chapter describes the details of a 12-bit, 75MS/s pipelined ADC prototype that was designed and implemented to evaluate the proposed calibration concept. In order to facilitate and expedite the evaluation, the chip was based on an existing, commercially available pipelined ADC in 0.35μm CMOS technology [82].

1.   ADC ARCHITECTURE

Figure 8-1 shows a block diagram of the experimental converter, which closely resembles the architecture of the original design before re-use. The pipeline core of this ADC is partitioned into a multi-bit first stage with an effective resolution of 3 bits, followed by eight stages, each resolving 1 bit effectively, and finally a 3-b flash sub-ADC.

Out of the 14 bits of raw data, the two least significant digits are used for calibration purposes only and truncated in the final conversion result. Stages 3-9 are implemented with 0.5-bit redundancy as standard 1.5-b stages (see e.g. [84]). As explained in section 3, the second stage of this design was modified to use one full bit of redundancy.

Compared to [82], the key modifications in the context of this work are the replacement of the stage 1 precision amplifier with an open-loop topology, and the addition of an off-chip digital post-processor to correct for resulting conversion errors. As discussed in chapter 6, the calibration could be extended to multiple open-loop stages in the converter front-end. For simplicity and improved transparency, only the first and most critical converter stage is converted to open-loop amplification in this demonstration vehicle.
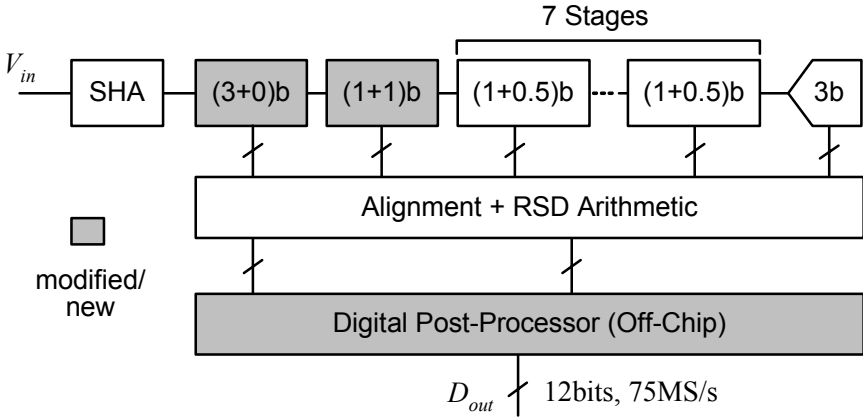
7 Stages

V_{in} SHA (3+0)b (1+1)b (1+0.5)b ---- (1+0.5)b 3b

Alignment + RSD Arithmetic

modified/
new

Digital Post-Processor (Off-Chip)

D_{out} 12bits, 75MS/s

Figure 8-1. Prototype architecture.

2.   STAGE 1

Figure 8-2 shows a schematic of the first converter stage. The sampling and DAC capacitor network of this circuit is identical to the implementation in [82], with the exception that here the sixteen poly-poly capacitors drive a resistively loaded open-loop amplifier with a nominal gain of 8. As in [82], a 4-bit flash converter is used to generate the coarse, local conversion result D. For this prototype, we chose a slightly different modulation scheme compared to the one discussed in chapter 7. The logic block between the sub-ADC and DAC switches implements the function

$$D_{mod} \quad D + [RNG \otimes \mathrm{mod}(D,2)], \qquad (8\text{-}1)$$

where $\otimes$ and *mod* denote the exclusive or and modulo operator respectively.

Figure 8-3 shows the stage's resulting residue plot for both random number generator states. While this alternative modulation achieves the same random switching between top/bottom segments, it has the advantage that each DAC state spans two segments in both modulation states of *RNG*. Therefore, the entire amplifier transfer function can be measured over a single, constant DAC code. This is advantageous for diagnostic purposes, since it provides a simple way to characterize the open-loop amplifier independent of potential DAC inaccuracy.
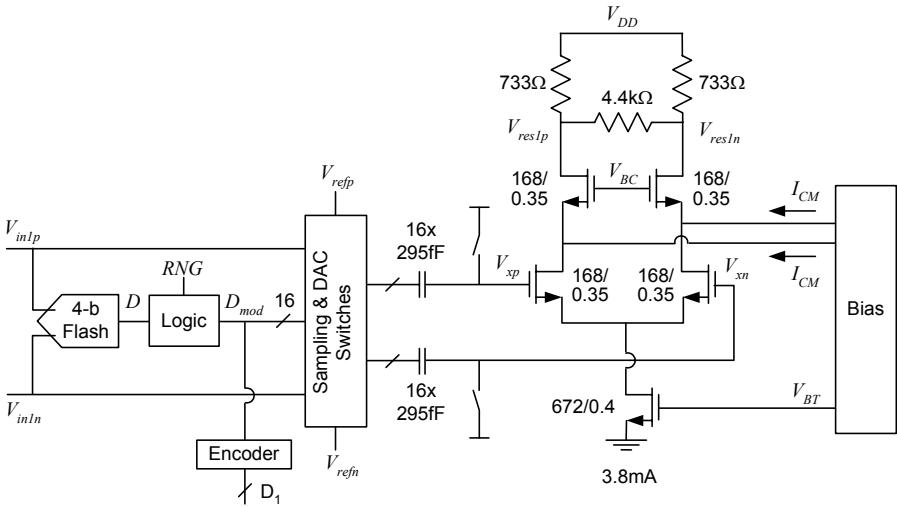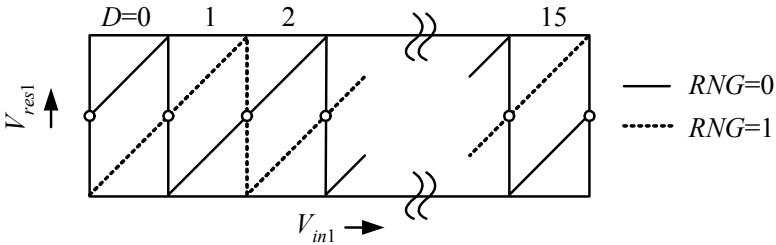
Figure 8-2. Stage 1 implementation.

Figure 8-3. Stage 1 residue plot.

2.1   Amplifier Design

In order to keep the 5th order distortion of the open-loop amplifier negligibly small, the quiescent point gate overdrive of the differential pair was chosen slightly larger than 250mV. With a reference voltage of 1V, the input swing of the differential is approximately 125mV. Therefore, the fractional swing α, given by V_{xmax}/V_{OV} is approximately 0.5. From Figure 5-6 we see that this choice results in a 5th order error of less than 0.1%, which corresponds to ½ LSB at the 9-bit backend resolution.

A π-load configuration was chosen for the amplifier to decouple the choice of common mode output level from differential gain requirements.

The equivalent, single ended Thévenin output resistance of this network is given by

$$R_{eq} \quad 733\Omega \,\|\, 2.2k\Omega \quad 550\Omega \tag{8-2}$$

For the given load conditions, this value was chosen to match the speed of the original, closed loop amplifier, which achieves settling to within $1/8^{th}$ LSB in 5ns. The total load capacitance of the amplifier is approximately 0.8pF, where 0.3pF are due to the sampling capacitors of the second converter stage, 0.35pF stem from parasitic junctions and the remaining portion is given by wiring capacitance. The total input referred noise contribution from this stage was found through simulation as approximately 50μVrms or equivalently 0.1LSBrms.

2.2 Biasing

A replica-biasing network controls both the amplifier tail current and common mode output level. Figure 8-4 illustrates the conceptual approach for the tail current generation. In this circuit, the output voltage of a scaled replica open-loop amplifier is forced to equal the reference voltage through negative feedback. Since the input of the replica is chosen $V_{ref}/8$, the gain of the stage is set to approximately 8. The resulting tail current is copied into the main amplifier to yield an equivalent gain factor in the signal path.
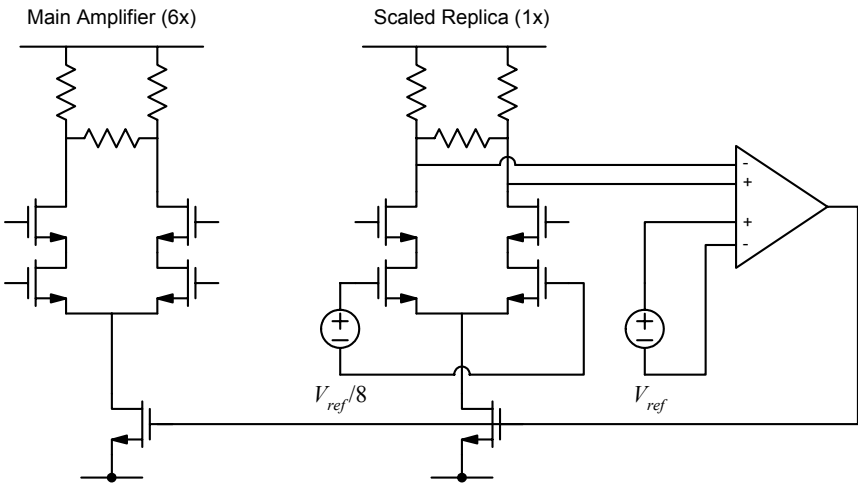

Figure 8-4. Replica biasing.

It should be noted that this replica technique is not very precise in terms of absolute accuracy. Since $V_{ref}/8 = 125$ mV, amplifier offset voltages on the order of 10mV can cause an error of 10% in the obtained gain factor. Nevertheless, there are two key benefits to the approach. First, the G_mR product is set at least reasonably close to the ideal value, which helps reduce the required range of the digital correction (p_1). Secondly, the replica technique decreases the temperature coefficient of the amplifier and thereby helps loosen the tracking requirements in the digital parameter estimation loops. The measurement results shown in chapter 9 illustrate this effect further.

A circuit similar to that in Figure 8-4 was used to control the common mode output level of the amplifier. The currents I_{CM} in Figure 8-2 are derived from a second replica feedback loop that sets the common mode output level to 2V.

2.3   Additional Desensitization

One potential problem in the implementation of the open-loop stage lies in the implicit assumption of the calibration algorithm that the amplifier coefficients are constant, and independent of signal activity.

First order calculations show that full scale current steering in the differential pair can result in a temperature difference of approximately 1 degree in it's two half circuits. With typical temperature coefficients in IC components of 0.1%/degree, this can result in an input referred 0.5LSB-conversion error. Since the minimum, local thermal time constants in a typical silicon substrate are on the order of 10μs [106], the temperature change cannot be tracked by the fairly slow digital calibration loops. In this implementation, we therefore used extensive device interleaving and n+ diffusion load resistors with low thermal resistance to mitigate the effect of signal dependent self-heating.

Another form of signal dependent coefficient modulation can occur through supply bounce or common mode variations. To address this issue, we included output cascodes (see Figure 8-2) to improve both the amplifier's power supply and input common rejection ratios. For instance, simulations show that the cascodes reduce common mode sensitivity significantly. Without cascodes, changes in the quiescent point drain-source voltages modulate the transconductance in the short channel differential pair devices. Simulation predicts a worst case input referred error of 1LSB for common mode variations of approximately 100mV. With cascodes, this effect is reduced by the intrinsic transistor gain, which is on the order of 30 in 0.35μm technology. A further improvement of the input common mode

rejection ratio was achieved by employing the replica tail current biasing approach proposed in [89].

   All of the above design choices can be regarded as fairly conservative. The primary objective here was to guarantee successful evaluation of the digital calibration concept. In future work, a more aggressive design approach that eliminates e.g. the additional cascodes could be considered.

3.   STAGE 2

   Since there is no redundancy in the first stage of this converter, any sub-ADC errors will cause its residue to exceed the $\pm V_{ref}$ full-scale bounds. In order to accommodate such errors, we modified the second stage of the original design [82] to achieve input overranging capability. Two extra comparators were added to the traditional 1.5-bit stage results to obtain the residue characteristic shown in Figure 8-5. With this arrangement, input over range of up to ½$Vref$ is mapped back to within the $\pm V_{ref}$ boundaries. With a nominal gain of 8 in stage 1 and V_{ref}=1V, this allows stage 1 comparator offsets of up to $\pm 0.5V/8 = \pm 62.5mV$.
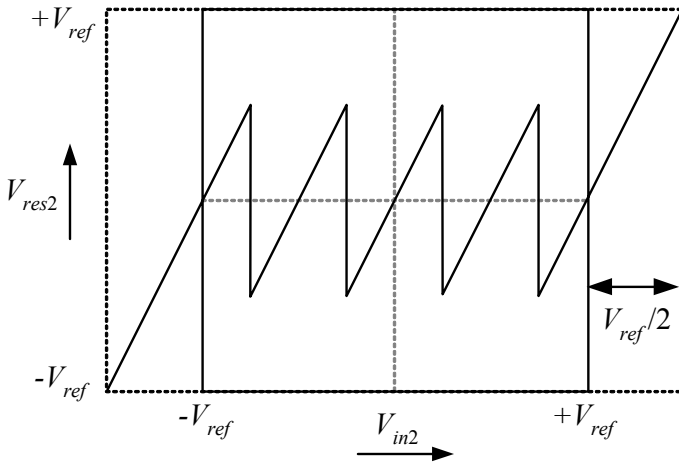

Figure 8-5. Stage 2 residue plot.

4. POST-PROCESSOR

A post-processor for digital correction and parameter estimation was implemented on an external field programmable gate array (FPGA). The FPGA was designed to perform the correction and parameter estimation in real time, at the full clock speed of 75MHz. Figure 8.6 details the interface between the prototype converter core and the FPGA post-processor.

All required arithmetic and logic functions, including a 64-stage pseudo random number generator that produces the *RNG* modulation signal were simulated and synthesized using the Verilog hardware description language.

Most of the post-processor's elements are generic adders, counters and registers. The required look-up table for cubic error correction was implemented using an incremental look-up table based on magnitude comparison (see Figure 8-7).

This circuit uses a bank of ROMs that generate digital thresholds for each LSB increment in the correction as a function of p_3. An advantage of this scheme is that the ROM look-up tables are not connected to the fast 75MHz signal path data. Instead, the only ROM input is parameter p_3, which changes at the slow update rate of f_s/N. This approach was necessary to achieve the desired throughput with an FPGA based design. The total ROM size was 64kBits for a reasonable p_3 range that covers temperature and process variations.

The hardware for the difference estimators D_1 and D_2 (see Figure 7-10) was implemented using dual port RAM macros in the FPGA. Each RAM word emulates an 8-bit counter bin that is incremented for a particular code hit in the top transfer curves (see Figure 7-8).
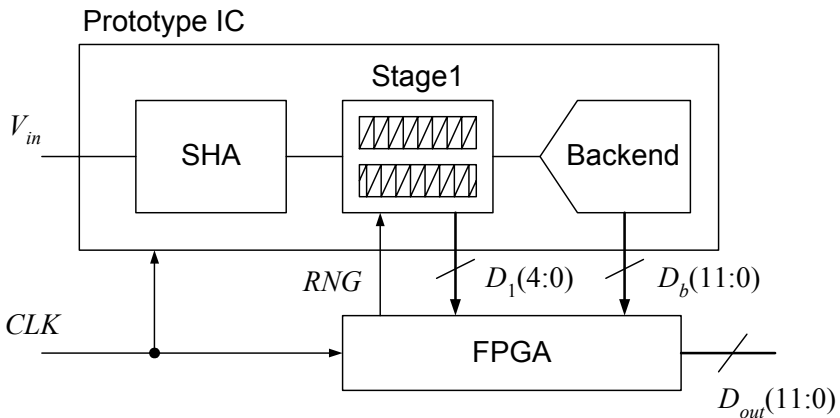
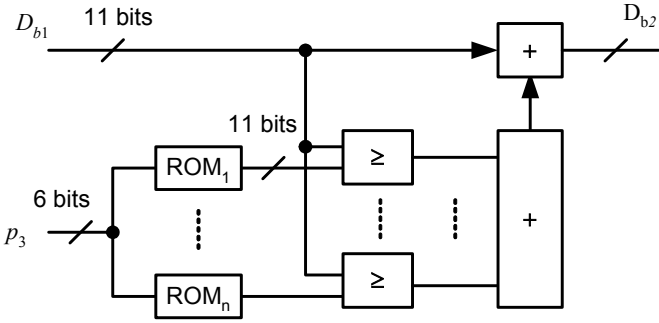

Figure 8-6. ADC-FPGA interface.

Figure 8-7. Incremental error look-up for cubic nonlinearity correction.

For each estimator (D_1, D_2), a histogram of 32 such bins is used. The histograms are periodically reset after N=50,000 samples. Before reset, the cumulative sum of the bins is used to find D_1 and D_2 as described in section 5 of chapter 7.

In order to reduce the total number of bins needed, the histograms were taken from a truncated 9-bit version of the backend data.

Chapter 9

EXPERIMENTAL RESULTS

1. LAYOUT AND PACKAGING

The prototype ADC described in the previous chapter was fabricated in a 0.35μm double-poly, quadruple metal (DPQM) CMOS process. A micrograph of the 7.9mm^2 chip is shown in Figure 9-1. Except for the re-design of stage 1 and the minor modifications in stage 2, the layout is largely unchanged from the original design [82].
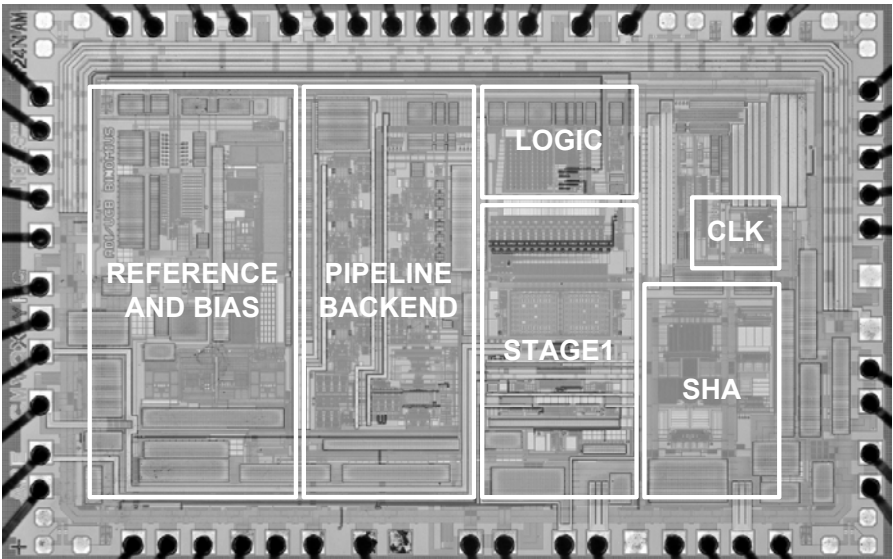
Figure 9-1. Die micrograph.

The available substrate in this fabrication run consisted of low resistance p+ with an epitaxial p- top layer. While epitaxial substrates provide good latch-up immunity, they have the disadvantage of creating low resistance paths for coupling undesired signals [107]. Therefore, special care was taken to yield a good conductive die attachment to the package lead frame which was connected to ground plane of the evaluation board through a low impedance path.

The chips were assembled in a 48-pin LQFP plastic package with 7mm x 7mm cavity. The bonding diagram and associated electrical pin-out are shown in Figure 9-2 and Table 9.1.

In order to investigate on the converter's temperature sensitivity, the two signals TEMPF (temperature force) and TEMPS (temperature sense) were added to the pin-out. These pins connect to an on-chip power transistor and a pn-junction that were placed near the open-loop amplifier of stage 1. These devices were used in the evaluation of the prototype to create and measure local operating temperature transients.
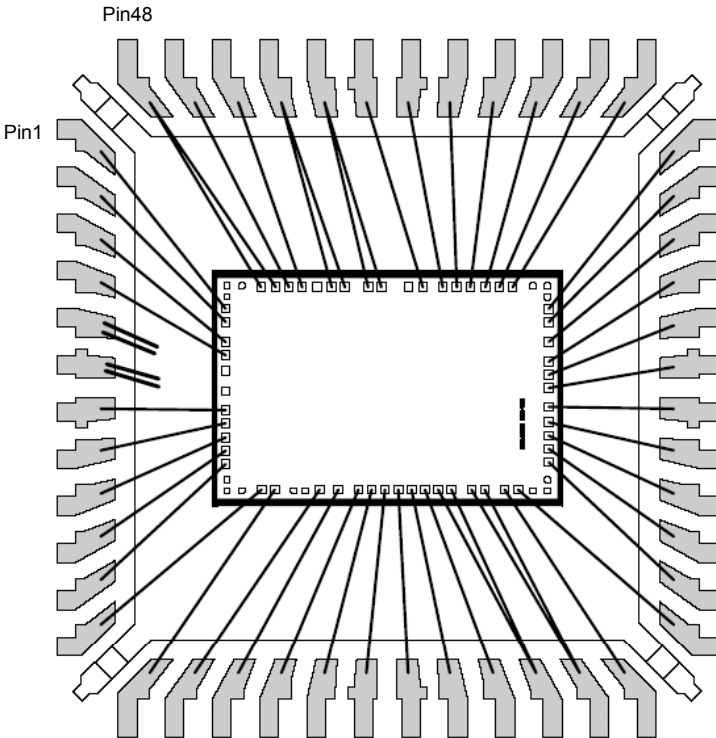

Figure 9-2. Bonding diagram.

Table 9-1. Pinout.

Pin	Name	Remark	Pin	Name	Remark
1	AGND	Analog Ground	25	$D_b[12]$	Backend Raw Data, MSB
2	AGND		26	$D_b[11]$	
3	AVDD	Analog Supply	27	$D_1[0]$	Stage1 Raw Data, LSB
4	AVDD		28	$D_1[1]$	
5			29	$D_1[2]$	
6			30	RNG	
7	CLK		31	AVDD	
8	PWDN	Power Down	32	AGND	
9	$D_1[3]$	Stage1 Raw Data, MSB	33	AGND	
10	$D_b[0]$	Backend Raw Data, LSB	34	AVDD	
11	$D_b[1]$		35	TEMPF	Power Transistor for Chip heating
12	$D_b[2]$		36	BMODE	Replica Bias On/Off
13	$D_b[3]$		37	VREF	Reference Input
14	DGND	Output Driver Ground	38	TWEAK	External Bias, used when BMODE=0
15	DVDD	Output Driver Supply	39	VREFN	Negative Reference Bypass
16	$D_b[4]$		40	VREFN	
17	$D_b[5]$		41	VREFP	Positive Reference Bypass
18	$D_b[6]$		42	VREFP	
19	$D_b[7]$		43	TEMPS	PN junction for Temperature Sensing
20	$D_b[8]$		44	AVDD	
21	$D_b[9]$		45	CML	Common Mode Reference Output
22	DGND		46	VINP	Positive Converter Input
23	DVDD		47	VINN	Negative Converter Input
24	$D_b[10]$		48	AVSS	

2.   TEST SETUP

The basic setup for the experimental testing is shown in Figure 9-3. For optimum performance, the packaged dice were soldered onto printed circuit boards that closely match the description in [108].

Both the system clock and ADC input signal were generated with high performance RF signal generators. For all sine wave tests, a band pass filter was used to reduce spurious components in the input signal.

The raw ADC output data is post-processed by the FPGA and subsequently captured in a FIFO at the true 75MHz clock speed. The stored data packets are then transferred at a lower clock rate to a personal computer via a data acquisition card.
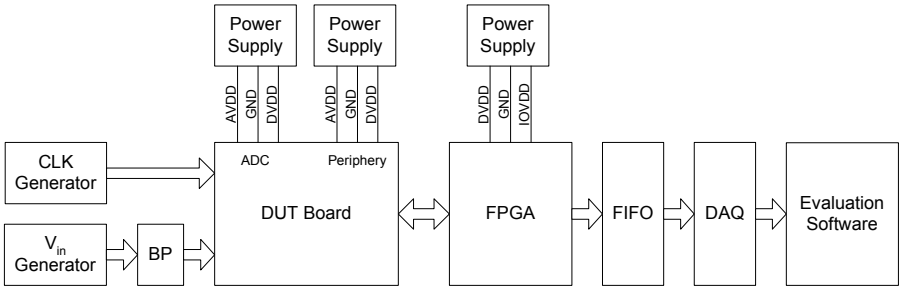
Figure 9-3. Test setup.

Table 9-2. Test equipment.

CLK Generator	Hewlett Packard 8644B
V_{in} Generator	Hewlett Packard 8644A
BP Filter	Allen Avionics F-series (e.g. F3962-20PO)
	K&L Tunable BP 5BT-30/76-5-N/N
DUT Board	Custom Design
Power Supplies	Agilent E3630A
FPGA Board	Xilinx HW-AFX-PQ240-100
	with XCV400E FPGA
FIFO	Analog Devices HSC-ADC-EVAL-SC
DAQ Card	National Instruments PCI-DIO-32
Evaluation Software	National Instruments LabView
	Matlab

3.   MEASURED RESULTS

3.1   Static Linearity

Figures 9-4 and 9-5 show the DNL and INL of the experimental converter without digital post-processing ($p_1=p_2=p_3=0$) for both *RNG* states. In both INL signatures, the gain compression of the open loop amplifier is clearly visible as a cubic bow. The transfer functions show several missing codes (DNL=-1) due to the inaccurate linear gain term and gain compression. Figure 9-6 shows the measured nonlinearity with active post-processing for linear and cubic errors only (p_1 and p_3). From this result, we see that the calibration removes all missing codes, and significantly improves the INL from its worst-case raw value of 18LSB to about 0.6LSB.
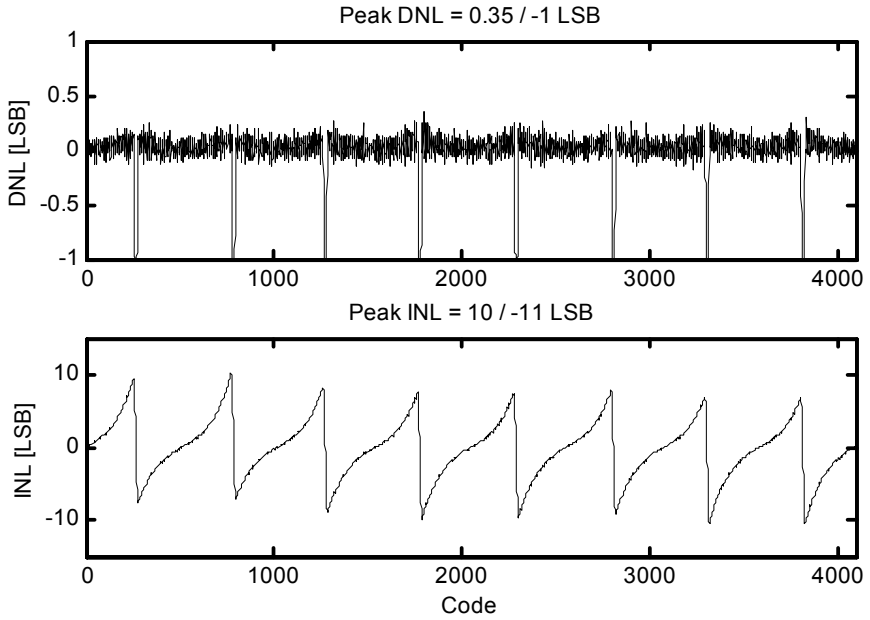
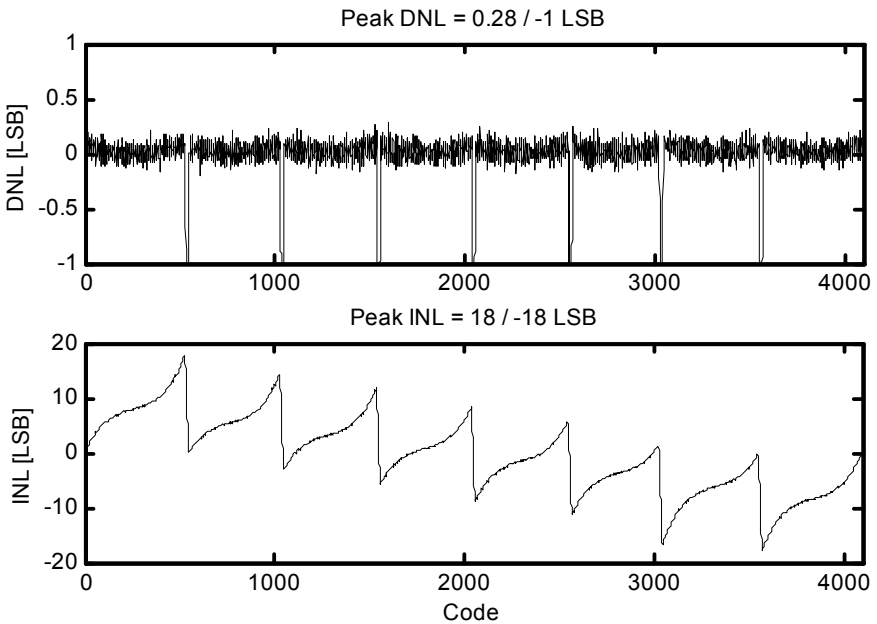*Figure 9-4.* Measured nonlinearity without calibration, *RNG*=0.

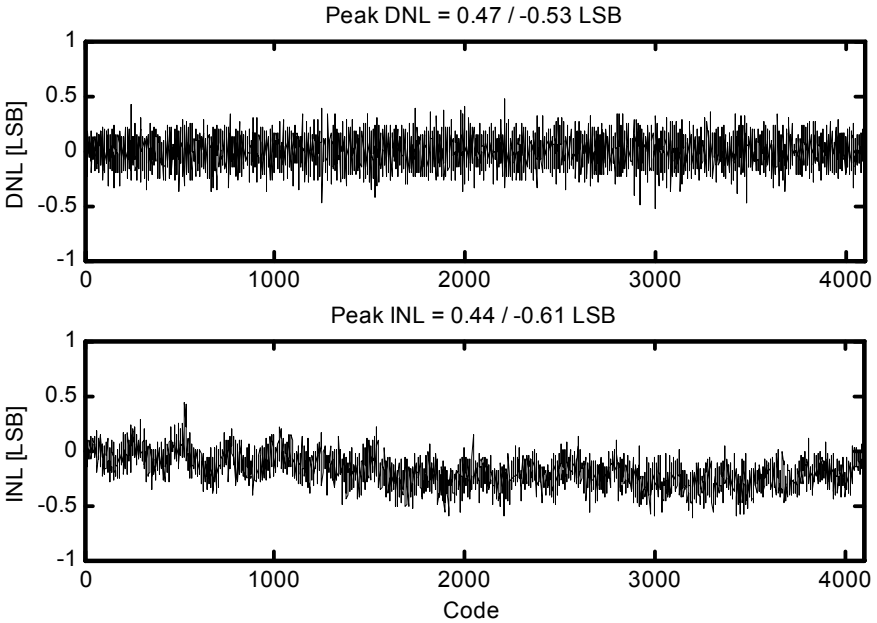*Figure 9-5.* Measured nonlinearity without calibration, *RNG*=1.

Figure 9-6. Measured nonlinearity with calibration.

The measured data showed that a quadratic correction using parameter p_2 was not necessary in this implementation. Figure 9-7 shows the measured peak INL versus the manually adjusted correction parameter p_2. From this graph we see that the optimum is close to p_2=0, i.e. no correction. From this result, we can conclude that there is sufficiently good matching (~better 0.3%) in the differential pair transistors and also only a small input referred offset in the converter backend. In Figure 9-7, with V_{ref}=1V, one backend LSB corresponds to $2V_{ref}/2^9\cong$4mV. The slight shift of the optimum region in Figure 9-7, may be due to a backend offset of about 1LSB or an equivalent voltage of 4mV.

Figure 9-8 shows the converter's INL with p_2 set to –7LSB. In this plot, we clearly see the quadratic error signature that stems from this maladjustment. For all further measurements, we eliminated the quadratic estimation loop and correction such that p_2=0.
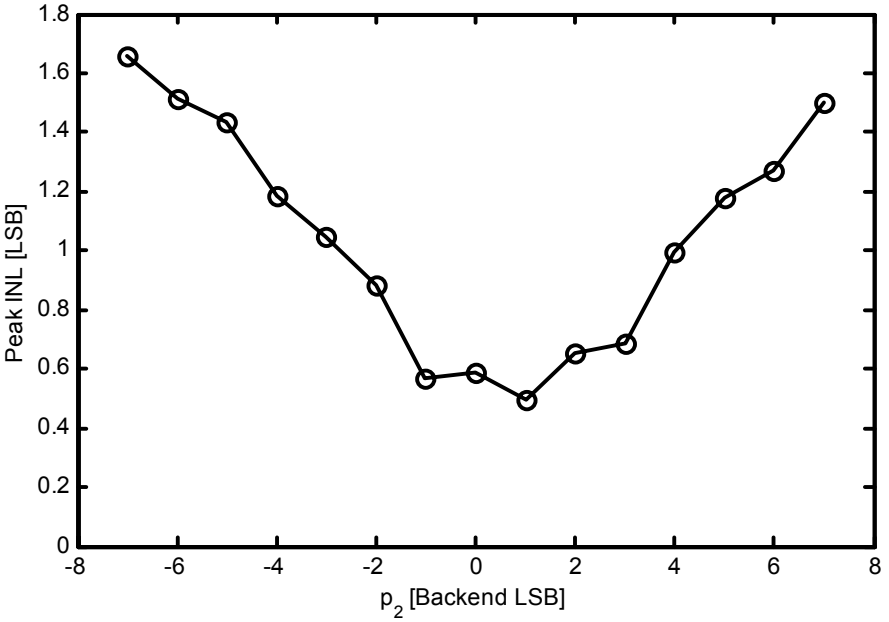
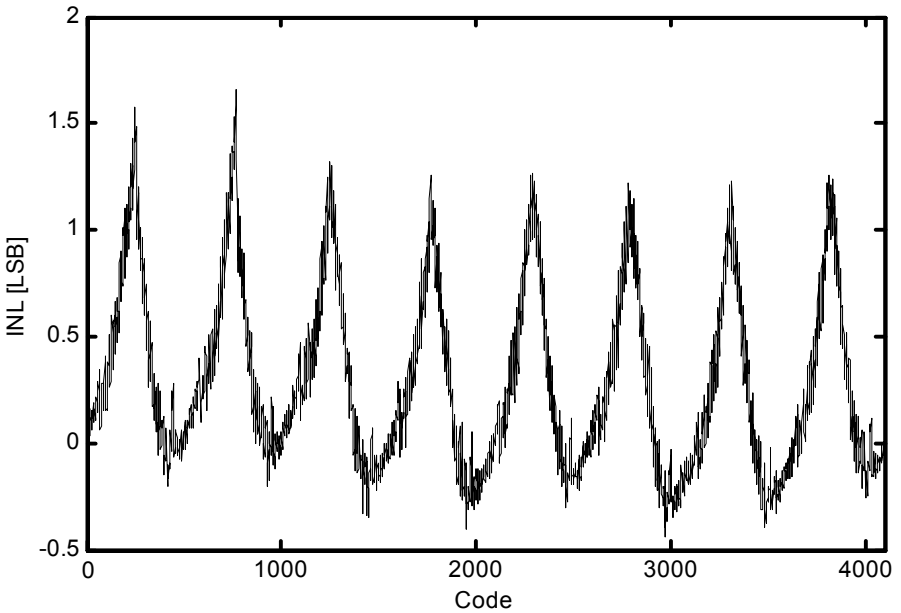*Figure 9-7.* Peak INL as a function of correction parameter p_2.

Figure 9-8. INL with p_2=-7 LSB.

3.2   **Dynamic Linearity and Noise Performance**

Shown in Figure 9-9 is the measured output spectrum at an input frequency of 40MHz. With calibration, spurious components are below -76dB. Figure 9-10 and 9-11 summarize the measured spectral performance as a function of sampling and input frequency. Due to the high performance front-end sample-and hold, the converter shows good performance beyond Nyquist input frequencies.
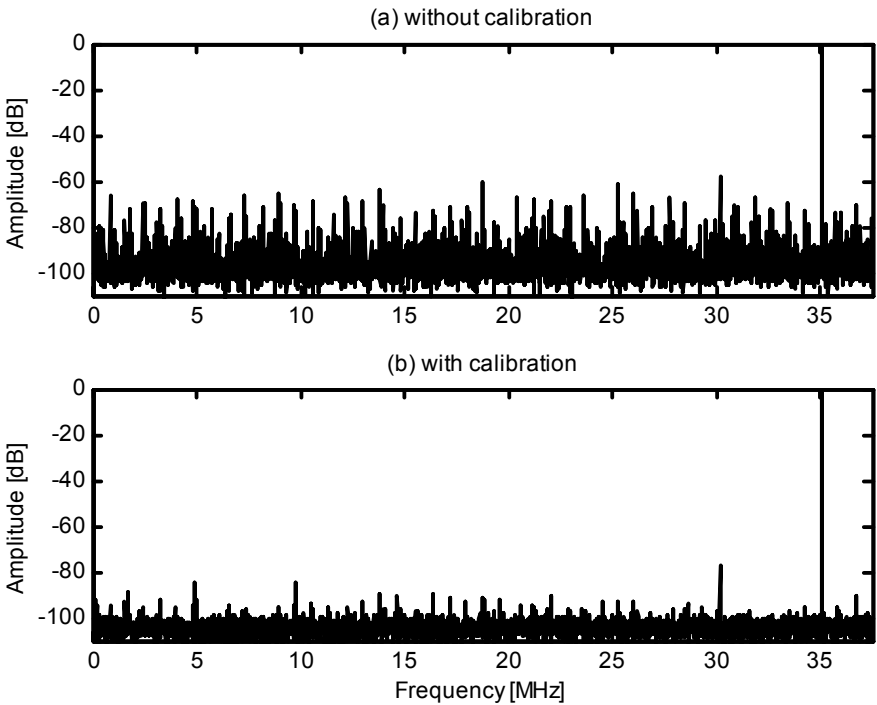

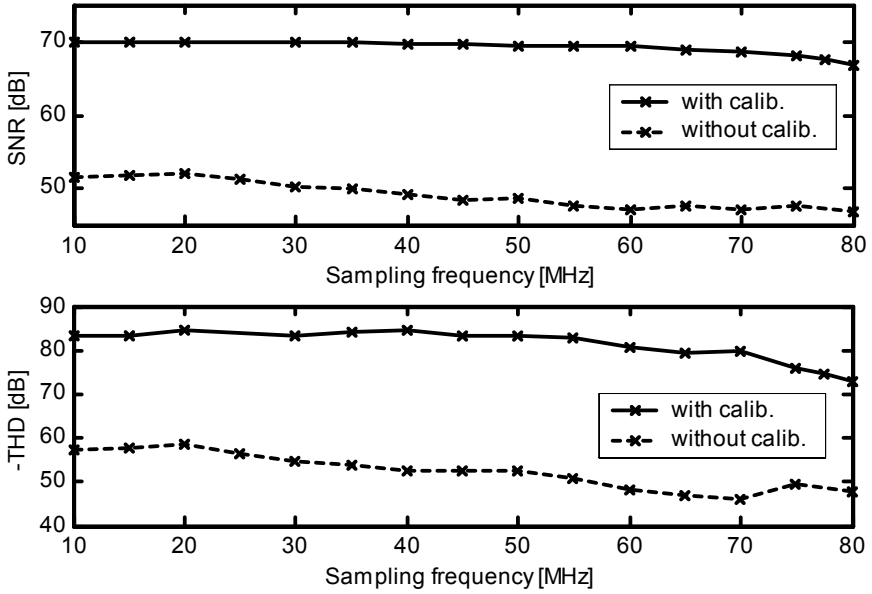*Figure 9-9.* Measured output spectrum (4096 point FFT).

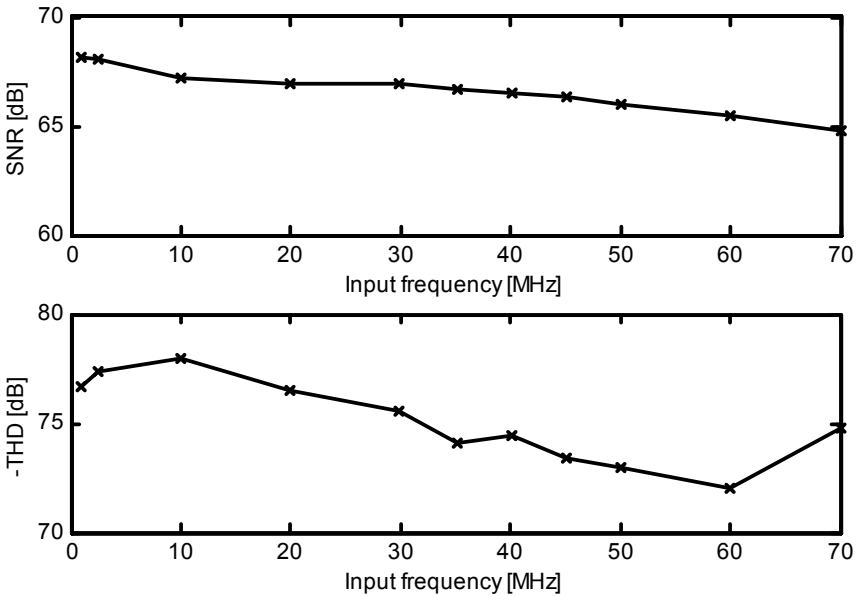*Figure 9-10.* Noise and distortion performance versus sampling frequency (f_{in}=1MHz).

Figure 9-11. Noise and distortion performance versus input frequency (f_s=75MHz).

3.3   **Temperature Tests**

In order to evaluate the robustness of the converter, we applied external temperature transients to its package using circuit cooler spray. Figure 9-12 illustrates the raw sensitivity of the system without any temperature compensating mechanisms.

In this measurement, the open loop amplifier was biased with a constant tail current and the adaptive LMS estimation loops were disabled ($\mu_1=\mu_3=0$) prior to the transient. The temperature variation is measured using a calibrated pn-junction (pin TEMPS) and the effective number of bits (ENOB) is continuously computed from the converter's response to a 1-MHz sine wave.

In contrast, Figure 9-13 shows the system response to a similar temperature pulse, but now with active LMS loops. The converter's ENOB and calibration parameter p_1 are plotted for two cases: (a) with constant tail current, and (b) with $Gm \cdot R$ replica biasing. In both cases, the ENOB remains relatively constant and exhibits mostly statistical ripple. With replica biasing, however, the tracking requirements on p_1 are reduced, resulting in a more robust overall system that can tolerate larger time constants in the LMS loops.
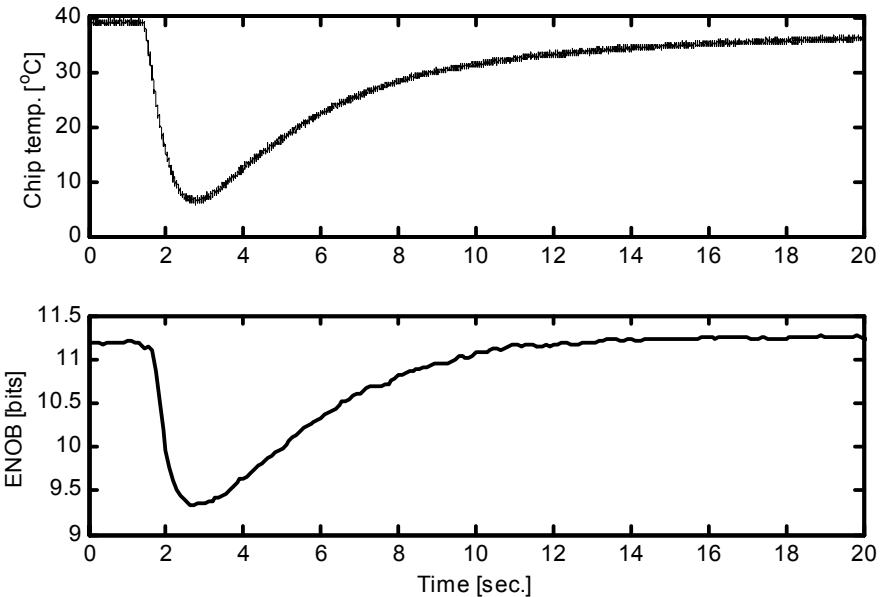

Figure 9-12. Measured temperature transient. Constant tail bias and LMS loops disabled ($\mu_1=\mu_3=0$).

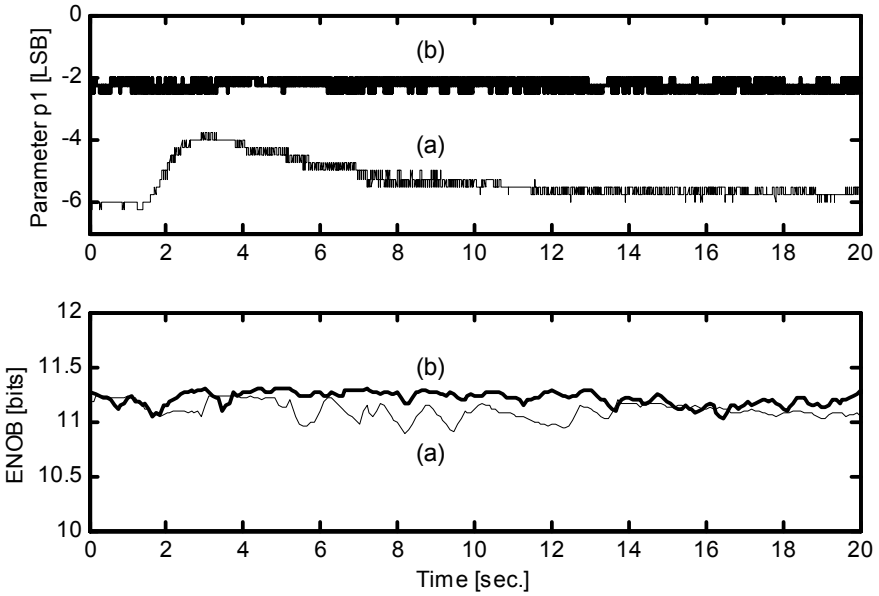*Figure 9-13.* Measured temperature transient with active LMS loops: (a) constant tail bias current, (b) with replica bias.

3.4   Power Reduction

Figure 9-14 compares the stage power breakdown in the original design [82] and this work. Pure transconductor power, accounting only for tail current invested to produce G_m, was reduced by 75% (34mW).
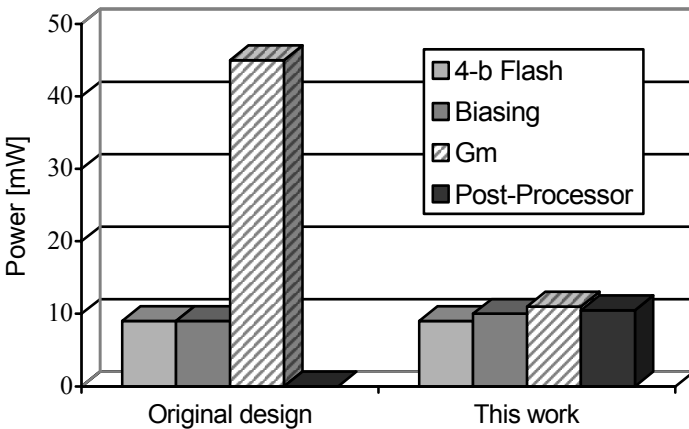

Figure 9-14. Stage 1 power breakdown.

Taking biasing networks into account, the overall amplifier power improved by 62% (33mW). Also shown in Figure 9-14 is the simulated power for digital post-processing. For the 0.35μm technology of this design, the simulated power consumption of the digital post-processor is only about one third of the power saved in the analog domain.

Since the first stage consumes only a fraction of the total power dissipation, the power savings with respect to the overall converter (340mW before re-design) are only about 11%. Figure 9-15 shows the achieved energy efficiency of *FOM2*=2.1pJ on the survey plot of chapter 2.

In an optimized, more aggressive deep sub-μm successor design, the expected benefit will increase for several reasons. First, multiple stages in the critical front-end of the converter could use open-loop amplification. Secondly, the gap between analog and digital power is expected to increase. As a result, there will be less power overhead due to digital calibration and also a potentially higher gain in analog efficiency due to improved compatibility with fine line technology.
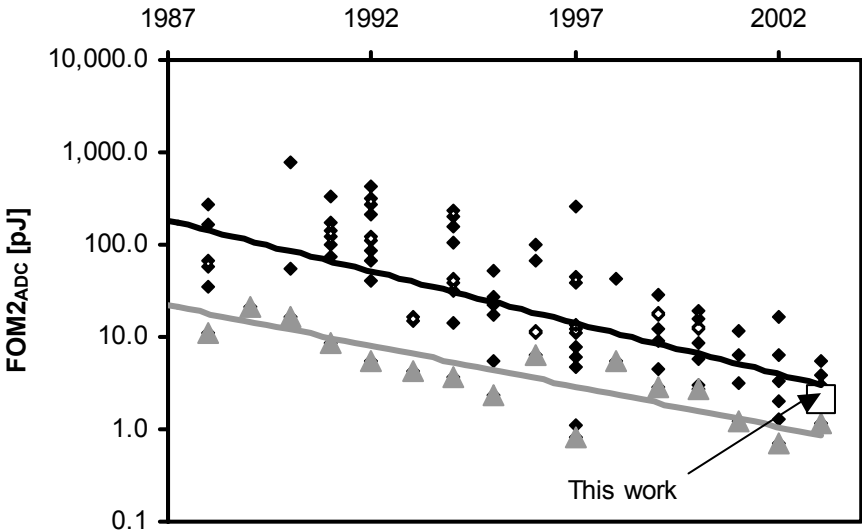

Figure 9-15. FOM2 performance of the prototype.

3.5 Performance Summary

Table 9.3 summarizes the performance of the experimental ADC.

Table 9-3. Performance summary (25°C).

Process, Area	0.35µm CMOS, 7.9mm2		
V_{DD}	3V		
Full Scale Range	2Vpp (differential)		
Resolution	12b		
Conv. Rate	75 MS/s		
	Without Post-Proc.	With Post-Proc.	
SNR	48dB	68.2dB 67dB	(fin=1MHz) (fin=40MHz)
THD	-50dB	-76dB -74dB	(fin=1MHz) (fin=40MHz)
SFDR	58dB	80dB 76dB	(fin=1MHz) (fin=40MHz)
DNL	-1, 0.35 LSB	-0.53, +0.47 LSB	
INL	-18, +18LSB	-0.61, +0.44 LSB	
PSRR (LF)	46dB		
Power: ADC Core Output Drivers	 290mW 24mW		

Several effects account for an INL that is larger than the predicted values of chapter 7. More detailed measurement results reveal that additional errors are mostly due to capacitor mismatch in the first two stages, uncompensated 5th order open-loop amplifier distortion and the onset of incomplete settling in the converter stages at f_s=75MHz.

4. POST-PROCESSOR COMPLEXITY

In order to investigate the post-processor's hardware complexity, we carried out synthesis and place & route design iterations using standard CMOS gate- and memory libraries. Excluding p_2, the digital logic for linear and cubic calibration can be implemented using 8400 gates, 64 bytes of RAM and 64 kBits of ROM. In 0.35µm CMOS technology, this translates to approximately $1.4mm^2$ of chip area, or approximately 18% of the prototype's area (see graphical illustration in Figure 9-16). Using 0.18µm-technology for comparison, the post-processor area decreases to $0.37mm^2$.
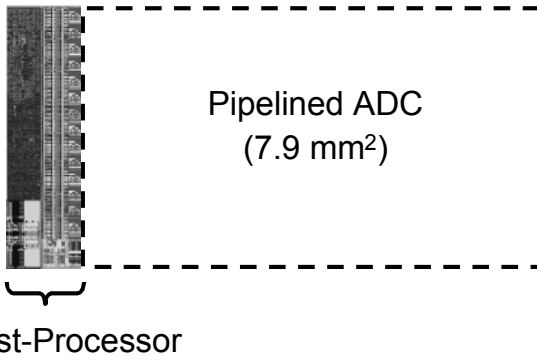
Figure 9-16. Estimated post-processor area for linear and cubic calibration.

Chapter 10

CONCLUSION

1.   SUMMARY

In the past decades, the continuing trend towards smaller feature sizes in integrated circuits has led to revolutionary progress in electronic systems. Since the beginning of this boom in the 1970's, the average price of a transistor has dropped from $1 to roughly 100 nano-cent [1]. Over the same period, we have seen nearly a 1-Million fold increase in computing power in digital microprocessors.

As we have shown in the surveys of chapter 2 and 3, smaller transistors have proven to be beneficial in the implementation of both analog and digital circuits. Yet, we are continuing to observe a large and growing gap in analog versus digital domain capabilities. Fundamentally, this trend is explained by the fact that some of the most severe analog circuit constraints do not scale with technology. For the most fundamental limitations such as noise and linearity, we are actually experiencing the onset of an inverse scaling trend. Excessive channel noise, decreasing supply voltages and intrinsic transistor gain tend to complicate the design of high dynamic rage, linear analog building blocks.

This book has explored the possibility to "digitally assist" an analog-to-digital converter, which can be regarded as one of the most basic and ubiquitous analog circuit. The proposed approach leverages the opportunity to treat analog-circuit nonlinearity as a digital-domain problem. With relaxed linearity specifications, analog circuits become simpler, faster and more power efficient.

The proposed digital nonlinearity compensation approach is applied to a pipelined ADC, whose most critical elements are the gain elements that

interface its individual stages. In the presented proof-of-concept prototype implementation, we show that significant power savings of up to 75% are possible when the conventional feedback amplifiers are replaced by simple open-loop gain stages.

The digital compensation approach constructed in chapters 6 and 7 is based on a digital pseudo random modulation that identifies and tracks amplifier nonlinearities in the background, allowing the system to track device and environmental variations without interrupting normal ADC operation. An important feature of the approach is that it does not introduce additional precision components or analog test signals. Therefore, calibration is achieved without sacrificing dynamic range or speed.

The measurement results documented in chapter 9 confirm the validity of the proposed scheme and highlight its potential. Particularly in fine line processes with low intrinsic device gain and limited supply headroom, the proposed scheme can be used to efficiently tradeoff analog precision for low power digital signal processing.

2.   SUGGESTIONS FOR FUTURE WORK

An obvious follow-up to the presented work is an extension of the technique to demonstrate an optimized deep sub-µm implementation with multi-stage calibration. Using multiple open-loop stages in the converter front-end will result in larger net power savings. At the same time, one could push more aggressively for higher conversion speed. In the presented proof-of-concept prototype, the conversion speed was limited by backend ADC stages that were re-used from a previous design.

Other opportunities exist in exploring similar calibration concepts for other ADC topologies. For instance, digital nonlinearity compensation could be used to remove distortion from folders in a folding ADC topology [109]. Mostly to remove the impact of nonlinearity, current folding-ADCs use analog interpolation networks that tend to limit their power efficiency and speed. Curing folder distortion in the digital domain is an opportunity to improve power and throughput of this particular converter topology.

A third, more aggressive vision, is to extend the digital correction in ADCs to include dynamic, frequency dependent errors. Dynamic error compensation has been discussed in literature [110, 111], but a feasible silicon implementation is yet to be demonstrated. The benefits of fully digital dynamic error compensation could be revolutionary, since for instance, complete settling would no longer be required in switched capacitor circuits.

More generally, a similar digital compensation approach to analog distortion could also be considered in sensors, transmission media and other physical domain elements that are limited by nonlinear effects.

This page intentionally left blank

Appendix A

OPEN-LOOP CHARGE REDISTRIBUTION

The proposed digital calibration technique assumes the presence of an ideal summing node in the open-loop pipeline stage. Equivalently, this requires that the pipeline stage has a transfer function of the form

$$V_{res} \quad b_1[V_{in} - V_{DAC}(D)] + b_2[V_{in} - V_{DAC}(D)]^2 + b_3[V_{in} - V_{DAC}(D)]^3 \ldots \quad (A-1)$$

In this family of power series, D is the local conversion result, and $V_{DAC}(D)$ represents the DAC-code dependent shift of each curve. In the following analysis, we show that V_{DAC} adds linearly to the input V_{in} of the open-loop charge redistribution network, despite the presence of nonlinear parasitics. For further discussion, consider the open-loop pipeline stage shown in Figure A-1.

If the capacitors $C_{1p} \ldots C_{jp}$ and $C_{1n} \ldots C_{jn}$ are sufficiently linear, the two arrays can be modeled by single Thévenin capacitors that are driven by equivalent DAC voltages in the redistribution phase. This is illustrated in Figure A-2. The respective equivalent values for the DAC voltages are

$$V_{dacp} \quad \frac{V_{refp}}{C_{SP}} \cdot \sum_{i\ 1}^{j} D_i \cdot C_{ip} + \frac{V_{refn}}{C_{SP}} \cdot \sum_{i\ 1}^{j} \overline{D}_i \cdot C_{ip}$$

$$V_{dacn} \quad \frac{V_{refn}}{C_{SN}} \cdot \sum_{i\ 1}^{j} D_i \cdot C_{in} + \frac{V_{refp}}{C_{SP}} \cdot \sum_{i\ 1}^{j} \overline{D}_i \cdot C_{in}$$

$$(A-2)$$
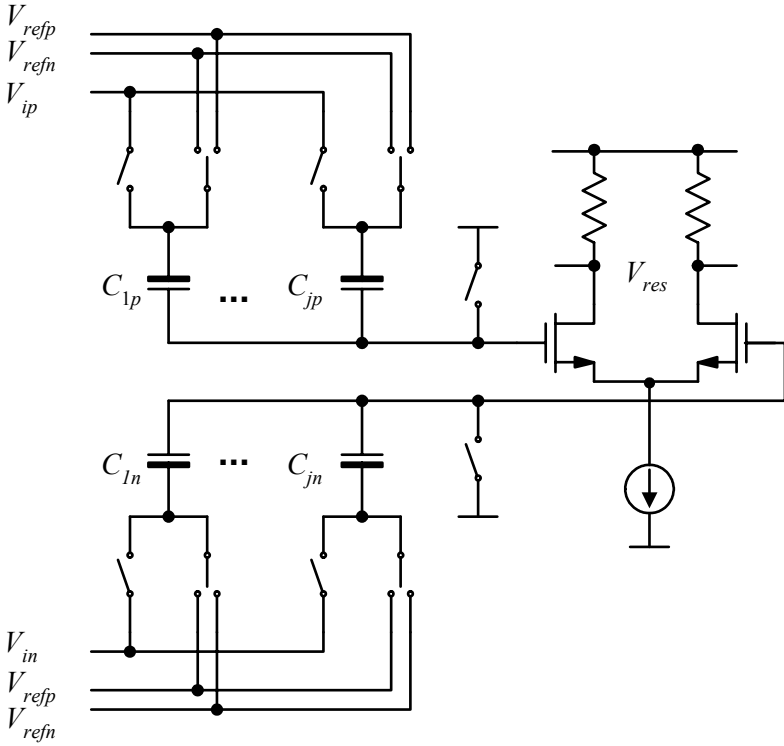
where D_i are the thermometer coded digital sub-DAC bits.
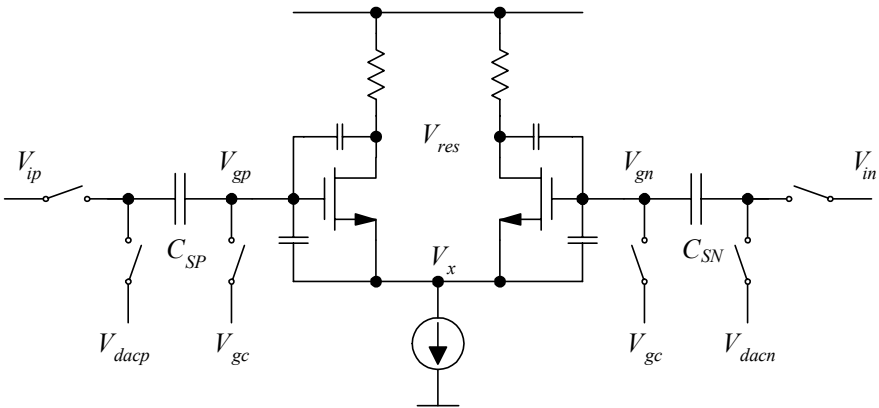
Figure A-1. Open-loop pipeline stage.

Figure A-2. Equivalent stage model.

The equivalent source capacitors in Figure A-2 are given by

$$C_{SP} \quad \sum_{i\ 1}^{j} C_{ip}$$
$$C_{SN} \quad \sum_{i\ 1}^{j} C_{in}$$

$$(A-3)$$

Also included in Figure A-2 are now the gate-source and gate-drain capacitors of the differential pair transistors. Looking into the two gates, we see a nonlinear capacitance, since: (1) C_{gs} is nonlinear, and (2) both C_{gs} and C_{gd} experience a nonlinear "Miller gain" [35] across them.

Finding a closed form expression for this nonlinear capacitance is fairly difficult and unnecessary. Instead, we consider here only the total charge on the gates, given by $Q_{gp}(V_{gp}, V_{gn})$ and $Q_{gn}(V_{gp}, V_{gn})$. These charges are completely determined by the gate voltages on each side. For ideal square law devices, these voltages are linked through the common source potential V_x by the expression

$$V_x \quad \frac{V_{gp} + V_{gn}}{2} - V_{TH} - \sqrt{V_{OV}^2 - \left(\frac{V_{gp} - V_{gn}}{2}\right)^2}$$

$$(A-4)$$

If we now write charge conservation equations for both clock phases and each gate node, we obtain

$$C_{SN}(V_{gc} - V_{in}) + Q_{gn}(V_{gc}, V_{gc})$$
$$C_{SN}(V_{gn} - V_{dacn}) + Q_{gn}(V_{gp}, V_{gn})$$

$$(A-5)$$

$$C_{SP}(V_{gc} - V_{ip}) + Q_{gp}(V_{gc}, V_{gc})$$
$$C_{SP}(V_{gp} - V_{dacp}) + Q_{gp}(V_{gp}, V_{gn})$$

$$(A-6)$$

In these expressions, V_{gn} and V_{gp} are the final voltages at the gates in the redistribution phase. Rearranging yields

$$V_{gn} \quad (V_{dacn} - V_{in}) + V_{gc} + \frac{Q_{gn}(V_{gc}, V_{gc})}{C_{SN}} - \frac{Q_{gn}(V_{gn}, V_{gp})}{C_{SN}} \qquad \text{(A-7)}$$

$$V_{gp} \quad (V_{dacp} - V_{ip}) + V_{gc} + \frac{Q_{gp}(V_{gc}, V_{gc})}{C_{SP}} - \frac{Q_{gp}(V_{gn}, V_{gp})}{C_{SP}}. \qquad \text{(A-8)}$$

From these equations, we see that the final gate voltages depend on the terms $(V_{dacn}\text{-}V_{in})$ and $(V_{dacp}\text{-}V_{ip})$. This means that independent of the nonlinear gate charges, the DAC voltages always add linearly to the input. Intuitively, this result is explained by the fact both the input and DAC operate on the same linear capacitor plates. If the exact charge relationships at the transistor gates were known, one could

– Subtract (A-7) from (A-8) to obtain differential variables V_{id} and V_{dacd}
– Approximate the gate charge relationships by a Taylor series
– Perform a power series reversion [92]

This procedure leads to an expression of the form

$$V_{gd} \quad c_1 \cdot (V_{dacd} - V_{id}) + c_2 \cdot (V_{dacd} - V_{id})^2 + c_3 \cdot (V_{dacd} - V_{id})^3 + ..., \qquad \text{(A-9)}$$

where V_{gd}, V_{dacd} and V_{id} are the differential gate, DAC and stage input voltages respectively. Next, we use the fact that that the differential pair output voltage can be expressed as

$$V_{res} \quad a_1 \cdot V_{gd} + a_2 \cdot V_{gd}^2 + a_3 \cdot V_{gd}^3 + ... \qquad \text{(A-10)}$$

From here, substitution of (A-9) into (A-10) leads to the desired result of (A-1).

Appendix B

ESTIMATOR VARIANCE

In the analysis below, we derive an approximate expression for the variance of the difference estimate D (equation (7-16)). For simplicity, we ignore the quantization error that stems from the discrete locations and boundaries of the cumulative histogram bins. This approximation is reasonable provided that the standard deviation of the distance estimate is larger than the histogram bin width, which corresponds to the LSB size of the backend converter (see considerations in section 6 of chapter 7).

Ignoring the quantization error, it follows from the setup of Figure 7-8 that

$$D \quad Y^*_{top} - y_{bot} \tag{B-1}$$

and thus

$$\text{var}(D) \quad \text{var}(Y^*_{top}). \tag{B-2}$$

In order to find the variance of the closest match Y^*_{top}, it is useful to partition the possible outcomes for each input sample $V_{in}(k)$ into three distinct events:

a) $RNG(k) = 1$, $V_{in}(k) \leq V$
b) $RNG(k) = 1$, $V_{in}(k) > V$
c) $RNG(k) = 0$

Provided that $V_{in}(k)$ is independent of $RNG(l)$ for all k and l, we can identify the following probabilities for each one of the above events

$$p_1 \quad 0.5 \cdot F(V)$$
$$p_2 \quad 0.5 \cdot (1 - F(V)) \tag{B-3}$$
$$p_3 \quad 0.5$$

where $F(V)$ denotes the cumulative distribution function of the samples $V_{in}(k)$, evaluated at $V_{in}=V$.

Now let the random variables N_1, N_2 and N_3 denote the number of occurrences for each possible event within a processing cycle of N samples. It then follows that these random variables have a multinomial distribution with parameters N and p_1, p_2 and p_3 respectively [112]. With respect to the setup in Fig. 7-8 we see that $N_1=CH(y_{bot})$ and that N_3 is the total number of samples that were processed using the upper transfer function segment (RNG=0).

After each processing cycle, the cumulative histogram bins are evaluated and we find Y_{top}^* such that its bin count is closest to the count in the reference bin y_{bot}, i.e.

$$Y_{top}^* \quad \arg\min_y \left(\left|CH(y_{bot}) - CH(y)\right|\right) \quad \arg\min_y \left(\left|N_1 - CH(y)\right|\right). \tag{B-4}$$

In the limit case of infinitely dense bins and a large number of samples, (B-4) is minimized such that $CH(Y_{top}^*)=N_1$ exactly. If we order all samples that make up $CH(Y_{top}^*)$, it follows that the largest one of these N_1 samples corresponds to the upper bin edge and consequently Y_{top}^* itself. Therefore, Y_{top}^* is given by the N_1^{th} order statistic in the sample of size N_3. Equivalently, Y_{top}^* represents the $(N_1/N_3)^{th}=P^{th}$ quantile of the samples processed by the upper transfer function segment (RNG=0).

Expressions for the variance of order statistics exist in literature, but they usually assume a fixed rank and sample size. Important to note in this analysis is that both the rank N_1 and sample size N_3 are random variables. A derivation from first principles that takes this randomness into account is desirable, but tends to yield complex results (see e.g. [113]). In the following steps, we use suitable simplifications to obtain an approximate, but sufficiently accurate result.

First, in order to relate the variance of Y_{top}^* to the statistics of V_{in} we can approximate for weakly nonlinear segment transfer functions

$$\mathrm{var}(Y_{top}^*) \quad \mathrm{var}(h_0(V^*)) \cong \mathrm{var}(b_1 V^*) \quad b_1^2 \, \mathrm{var}(V^*) \tag{B-5}$$

where V^* is the Pth quantile of an input sample of size N_3, and $P=N_1/N_3$. By conditioning on P we can rewrite

$$\text{var}(V^*) \quad \text{var}(E(V^*|P)) + E(\text{var}(V^*|P)) \tag{B-6}$$

For a uniform input distribution ($f(V_{in}(k))=1/\Delta$, $F(V_{in}(k))=V_{in}(k)/\Delta$), the conditional expectation of V^* in the first term of (B-6) is simply $F^{-1}(P)=P\cdot\Delta$. To capture a more general case, we use a linear gradient approximation for $F(V_{in})$ in the small region of interest around the estimation site V

$$F(V_{in}(k)) \cong F(V) + f(V) \cdot (V_{in}(k) - V) \tag{B-7}$$

Inverting this expression gives the approximate location of the quantile

$$F^{-1}(P) \quad \frac{P - F(V)}{f(V)} + V \tag{B-8}$$

The first term of (B-6) then becomes

$$\text{var}(E(V^*|P)) \quad \text{var}\left(\frac{P - F(V)}{f(V)} + V \right)$$

$$\frac{1}{f(V)^2} \text{var}(P) \tag{B-9}$$

$$\frac{1}{f(V)^2} \text{var}\left(\frac{N_1}{N_3} \right)$$

It is not straightforward to derive an exact expression for the variance of the quotient N_1/N_3. However, it is possible to obtain a good approximation through a second order Taylor expansion of the quotient [114]. This approximation is given by

$$\text{var}\left(\frac{N_1}{N_3} \right) \cong \left(\frac{E(N_1)}{E(N_3)} \right)^2 \cdot \left[\frac{\text{var}(N_1)}{E(N_1)^2} + \frac{\text{var}(N_3)}{E(N_3)^2} - \frac{2 \cdot \text{cov}(N_1, N_3)}{E(N_1) \cdot E(N_3)} \right] \tag{B-10}$$

Using formulae for the moments of the multinomial distribution of N_1 and N_3, we identify

$$E(N_1) \quad N \cdot p_1 \qquad \text{var}(N_1) \quad N \cdot p_1(1\text{-}p_1)$$

$$E(N_3) \quad N \cdot p_3 \qquad \text{var}(N_3) \quad N \cdot p_3(1\text{-}p_3) \tag{B-11}$$

$$\text{cov}(N_1, N_3) \quad -N \cdot p_1 \cdot p_3$$

Using (B-3) and substituting the above into (B-10) and (B-9) yields

$$\text{var}(E(V*|P)) \quad \frac{2 \cdot F(V) \cdot (1 + F(V))}{N \cdot f(V)^2}. \tag{B-12}$$

Next, consider the second term in (B-6). If $F(V_{in})$ is strictly increasing and continuous, a general approximation formula exists for the variance of a pth quantile Q in a sample of size M, with local density $f(Q)$ [114]

$$\text{var}(Q) \cong \frac{p(1-p)}{M \cdot f(Q)^2}. \tag{B-13}$$

Using this result, and noting that the sample size under consideration corresponds to N_3, we obtain for the second variance component of (B-6)

$$E(\text{var}(V*|P)) \cong \frac{1}{f(V)^2} E\left(\frac{P(1-P)}{N_3}\right) \quad \frac{1}{f(V)^2} E\left(\frac{\dfrac{N_1}{N_3}\left(1 - \dfrac{N_1}{N_3}\right)}{N_3}\right). \tag{B-14}$$

In general, the expected value of a function of random variables can be approximated through a Taylor series expansion of the form

$$E(g(A,B)) \cong g(E(A), E(B)) + \frac{1}{2}\frac{\partial^2 g}{\partial A^2}\text{var}(A)$$

$$+ \frac{1}{2}\frac{\partial^2 g}{\partial B^2}\text{var}(B) + \frac{1}{2}\frac{\partial^2 g}{\partial A \partial B}\text{cov}(A,B) \tag{B-15}$$

Detailed analysis shows that the second order terms in this approximation are negligibly small for the function inside the expected value operator of (B-14). Consequently, we can approximate using only the first term in (B-15), i.e.

$$E\left(\frac{\dfrac{N_1}{N_3}\left(1-\dfrac{N_1}{N_3}\right)}{N_3}\right) \cong \frac{\dfrac{E(N_1)}{E(N_3)}\left(1-\dfrac{E(N_1)}{E(N_3)}\right)}{E(N_3)}. \tag{B-16}$$

Using this simplification, and substituting (B-11) and (B-3) yields

$$E(\mathrm{var}(V^* \mid P)) \cong \frac{2 \cdot F(V) \cdot (1 - F(V))}{N \cdot f(V)^2}. \tag{B-17}$$

Finally, adding (B-12) and (B-17) yields the desired end result

$$\mathrm{var}(D) \cong \frac{4}{N} \cdot \frac{F(V)}{f(V)^2}. \tag{B-18}$$

Figure B-1 shows a simulation result that shows good agreement with the approximation of (B-18). In this example, a 100-run Monte Carlo simulation was performed for each value of V/Δ. The samples $V_{in}(k)$ have a Gaussian distribution with mean $\Delta/2$ and standard deviation $\Delta/3$. N=100,000 samples are collected in each Monte Carlo run until histogram evaluation.
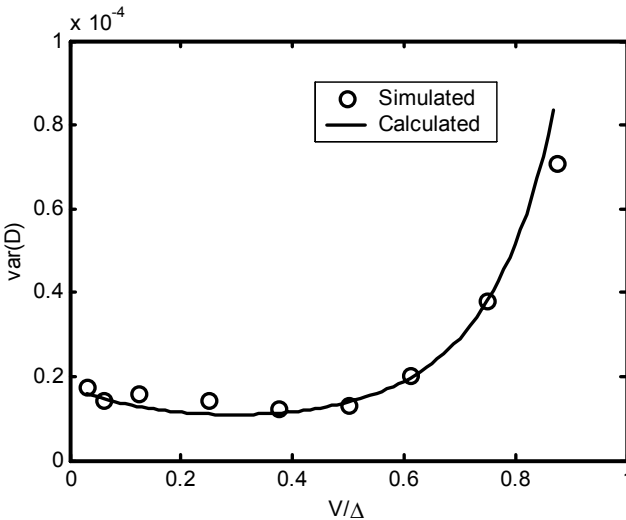

Figure B-1. Simulated estimator variance for Gaussian input.

This page intentionally left blank

Appendix C

LMS LOOP ANALYSIS

1. TIME CONSTANT

In the following analysis, we derive expressions for the tracking time constant and variance at the outputs (p_i) of the LMS loops in Figure 7-10. Figure C-1 shows a suitable block diagram for further consideration. The difference equation for this model is given by

$$p_i(n) \quad p_i(n-1) + \mu_i \left[\varepsilon_i(n-1) - \delta_i p_i(n-1) \right], \tag{C-1}$$

where n represents the index of the discrete time samples in the loop. Since the LMS loop is only updated every N samples, n relates to the discrete time index k of the converter's input samples as

$$n \quad N \cdot k . \tag{C-2}$$

In order to derive the envelope time constant of this loop without the estimator noise present, we let $\varepsilon_i = 0$.
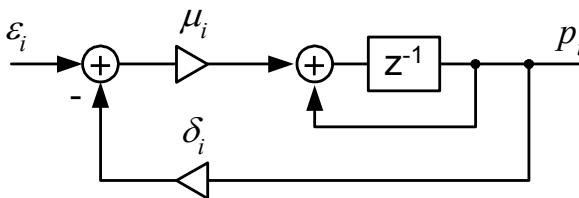

Figure C-1. LMS loop block diagram.

With this condition, (C-1) simplifies to

$$p_i(n) \quad p_i(n-1)\left[1 - \mu_i \delta_i\right]. \tag{C-3}$$

Assuming an initial condition $p_i(0)$, the parameter values at an arbitrary time index are given by

$$p_i(n) \quad p_i(0)\left[1 - \mu_i \delta_i\right]^n. \tag{C-4}$$

The time constant of the system is given by the time at which the initial condition has decayed to a value of $1/e$, hence

$$\frac{1}{e} p_i(0) \quad p_i(0)\left[1 - \mu_i \delta_i\right]^{\tau_n}. \tag{C-5}$$

Taking the natural logarithm on both sides and using the first order expansion

$$\ln(1 - x) \cong -x, \tag{C-6}$$

we obtain

$$\tau_n \cong \frac{1}{\mu_i \delta_i}. \tag{C-7}$$

This discrete time constant can be expressed in terms of absolute time using (C-2) and the fact that the converter samples the input every $1/f_s$ seconds. Therefore

$$\tau \cong \frac{N}{f_s} \frac{1}{\mu_i \delta_i}. \tag{C-8}$$

2.   OUTPUT VARIANCE

Next, we establish an expression for the variance in $p_i(n)$ given a certain variance in $\varepsilon_i(n)$ that is due to uncertainty in the difference estimators in Figure 7-8. From the recursion of (C-1) we obtain

$$p_i(n) \quad p_i(0)[1 - \mu_i \delta_i]^n + \mu_i \sum_{j\ 0}^{n-1} \varepsilon_i(j)[1 - \mu_i \delta_i]^{n-1-j} . \qquad (C\text{-}9)$$

Taking the variance of both sides, assuming that the $\varepsilon_i(j)$ terms are statistically independent and have equal and constant variance $var(\varepsilon_i)$, we obtain

$$\mathrm{var}[p_i(n)] \quad \mu_i^2 \, \mathrm{var}(\varepsilon_i) \sum_{j\ 0}^{n-1} [1 - \mu_i \delta_i]^{n-1-j} . \qquad (C\text{-}10)$$

For $n \to \infty$, which corresponds to steady state operation, we can use

$$\sum_{j\ 0}^{\infty} q^j \quad \frac{1}{1-q}; \qquad q < 1 \qquad (C\text{-}11)$$

to obtain

$$\mathrm{var}[p_i(n)] \quad \mu_i^2 \, \mathrm{var}(\varepsilon_i) \frac{1}{1 - (1 - \mu_i \delta_i)^2} . \qquad (C\text{-}12)$$

For the practical case of $\mu_i \delta_i \ll 1$, this expression is well approximated by

$$\mathrm{var}[p_i(n)] \cong \frac{1}{2} \frac{\mu_i}{\delta_i} \mathrm{var}(\varepsilon_i) \qquad (C\text{-}13)$$

3.   MAXIMUM GAIN PARAMETERS

In this section we establish an upper bound for the LMS loop gain parameters μ_i based on accuracy requirements in the ADC. Consider first the linear calibration loop of Figure 7-10. The sensitivity of the converter output D_{out} to variations in p_1 is given by

$$\frac{dD_{out}}{dp_1} \quad -D\,, \tag{C-14}$$

where D is the local sub-conversion result. Hence, the conversion result is most sensitive to variation in p_1 near the full-scale values of D. One way to establish an upper bound for the variance in p_1 is to model D as a random process and to find the resulting net noise in D_{out} through the product of the random variables D and p_1. In this discussion, we consider the worst-case DNL error of the transfer function instead.

For each increment in D (step size Δ in equation (7-8)), the error in D_{out} must be bounded to a fraction of an LSB. This requirement translates into

$$\mathrm{var}(p_1) \cdot \Delta^2 \leq L_1^2 \cdot \left(\frac{2}{2^{B_{tot}}}\right), \tag{C-15}$$

where B_{tot} corresponds to the overall resolution of the ADC, and L_1 is the allowable worst case DNL error in LSBrms due to variance in p_1. Using equation (7-8), and noting that $\mathrm{var}(\varepsilon_1)=\mathrm{var}(D_1)$ and $\delta_1=1$, this result modifies to

$$\mu_1 \leq 2L_1^2 \cdot \left(\frac{2}{2^{B_b}}\right)\frac{1}{\mathrm{var}(D_1)} \tag{C-16}$$

Analogous considerations based on worst-case DNL errors lead to similar equations for the quadratic and cubic calibration loops

$$\mu_2 \leq 2\delta_2 L_2^2 \cdot \left(\frac{2}{2^{B_b}}\right)\frac{1}{\mathrm{var}(E_1 - E_2)}$$

$$\mu_3 \leq 2\delta_3 L_3^2 \cdot \left(\frac{2}{2^{B_b}}\right)\frac{1}{\mathrm{var}(D_1 - D_2)} \tag{C-17}$$

The variance terms in the above equations can be found using (7-17). As required by the algorithm, the estimate D_1 is taken as close as possible to the segment edge, i.e. $V=V_{min}$ in (7-17). Therefore

$$\text{var}(D_1) \quad \frac{4}{N}\frac{V_{\min}}{\Delta} \qquad\qquad\text{(C-18)}$$

This result also holds for E_1 and E_2 which should also be taken as close as possible to the segment edge for maximum loop sensitivity (see equation (7-13)). Neglecting correlation in the two difference estimates, we therefore have

$$\text{var}(E_1 - E_2) \cong \text{var}(E_1) + \text{var}(E_2) \quad \frac{8}{N}\frac{V_{\min}}{\Delta} \qquad\qquad\text{(C-19)}$$

As required by the algorithm, the estimate D_2 should be taken close to the segment center $\Delta/2$. As a result, we obtain

$$\text{var}(D_1 - D_2) \cong \text{var}(D_1) + \text{var}(D_2)$$

$$\cong \frac{4}{N}\frac{V_{\min}}{\Delta} + \frac{4}{N}\frac{\Delta/2}{\Delta} \quad \frac{2}{N}\left(1 + \frac{V_{\min}}{\Delta}\right) \qquad\qquad\text{(C-20)}$$

Combining (C-16)-(C-20) leads to the final result stated in equations (7-22) and (7-23).

This page intentionally left blank

References

[1]   G. Moore, "No Exponential is Forever: But 'Forever' can be delayed!," *ISSCC Dig. Techn. Papers*, pp. 21-23, Feb. 2003.

[2]   A. M. Abo, *Design for reliability of low-voltage, switched-capacitor circuits*: PhD Thesis, University of California, Berkeley, 1999.

[3]   S. H. Lewis, et al., "A 10-b 20-Msample/s analog-to-digital converter," *IEEE J. of Solid-State Circuits*, pp. 351-358, Mar. 1992.

[4]   A. N. Karanicolas, et al., "A 15-b 1-Msample/s digitally self-calibrated pipeline ADC," *IEEE J. of Solid-State Circuits*, pp. 1207-1215, Dec. 1993.

[5]   B. Murmann and B. E. Boser, "A 12-bit 75-MS/s Pipelined ADC using Open-Loop Residue Amplification," *IEEE J. Solid-State Circuits*, pp. 2040-2050, Dec. 2003.

[6]   Intel, (2003) "Moore's Law," [Online]. Available: http://www.intel.com/research/silicon/mooreslaw.htm.

[7]   Semiconductor Industry Association (SIA), (2003) "International Technology Roadmap - 2002 Update," [Online]. Available: http://public.itrs.net/.

[8]   D. Harris et. al, (2003) "The Fanout-of-4 Inverter Delay Metric," [Online]. Available: http://www3.hmc.edu/~harris/research/FO4.pdf.

[9]   M. A. Horowitz, "Circuits and Interconnects In Aggressively Scaled CMOS," *International Symposium on Computer Architecture*, June 2000.

[10]   J. Stokes, (2003) " Behind the benchmarks: SPEC, GFLOPS, MIPS et al.," [Online]. Available: http://arstechnica.com/cpu/2q99/benchmarking-1.html.

[11]   J. Stinson, *EE371 Course Notes, Stanford University*, 2003.

[12]   S. Borkar, "Design challenges of technology scaling," *IEEE Micro*, pp. 23-29, Apr. 1999.

[13]   K. Poulton, et al., "A 20 GS/s 8b ADC with a 1MB Memory in 0.18µm CMOS," *ISSCC Dig. Techn. Papers*, pp. 318-319, Feb. 2003.

[14]   A. Behzad, et al., "Direct Conversion CMOS Transceiver with Automatic Frequency Control for 802.11a Wireless LANs," *ISSCC Dig. Techn. Papers*, pp. 356-356, Feb. 2003.

[15]   I. Bouras, et al., "A Digitally Calibrated 5.15 - 5.825GHz Transceiver for 802.11a Wireless LANs in 0.18um CMOS," *ISSCC Dig. Techn. Papers*, pp. 352-353, Feb. 2003.

[16]   P. Zhang, et al., "A Direct Conversion CMOS Transceiver for IEEE 802.11 WLANs," *ISSCC Dig. Techn. Papers*, pp. 354-355, Feb. 2003.

[17]   R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journal on Selected Areas in Communications*, pp. 539-50, Apr. 1999.

[18]   R. v. d. Plassche, *CMOS Integrated Analog-to-Digital and Digital-to-Analog Converters*, 2nd ed. Boston: Kluwer Academic Publishers, 2003.

[19]   F. Goodenough, "Analog technology of all varieties dominate ISSCC," *Electronic Design*, pp. 96, Feb. 19, 1996.

[20]   M. Vogels and G. Gielen, "Architectural selection of A/D converters," *Proc. Design Automation Conference*, pp. 974-977, 2003.

[21]   J. Thomson, et al., "An integrated 802.11a baseband and MAC processor," *ISSCC Dig. Techn. Papers*, pp. 126-127, Feb. 2002.

[22]   S. M. Jamal, et al., "A 10-b 120-Msample/s time-interleaved analog-to-digital converter with digital background calibration," *IEEE J. of Solid-State Circuits*, pp. 1618-1627, Dec. 2002.

[23]   P. C. Yu, et al., "A 14 b 40 MSample/s pipelined ADC with DFCA," *ISSCC Dig. Techn. Papers*, pp. 136-137, Feb. 2001.

[24]   J. Elbornsson, "Blind estimation and error correction in a CMOS ADC," *Proc. ASIC/SOC Conference*, pp. 124-128, Sept. 2000.

[25]   E. B. Blecker, et al., "Digital background calibration of an algorithmic analog-to-digital converter using a simplified queue," *IEEE J. of Solid-State Circuits*, pp. 1059-1062, June 2003.

[26]   I. Galton, "Digital cancellation of D/A converter noise in pipelined A/D converters," *IEEE Trans. Ckts. Syst II*, pp. 185-196, March 2000.

[27]   J. Ming and S. H. Lewis, "An 8-bit 80-Msample/s pipelined analog-to-digital converter with background calibration," *IEEE J. of Solid-State Circuits*, pp. 1489-1497, Oct. 2001.

[28]   J. Li and U.-K. Moon, "Background calibration techniques for multistage pipelined ADCs with digital redundancy," *IEEE Trans. Ckts. Syst. II*, pp. 531-538, Sept. 2003.

[29]   A.-J. Annema, "Analog circuit performance and process scaling," *IEEE Trans. Ckts. Syst. II*, pp. 711-725, June 1999.

[30]   K. Bult, "Broadband communication circuits in pure digital deep sub-micron CMOS," *ISSCC Dig. Techn. Papers*, pp. 76-77, Feb. 1999.

[31]   J. M. Rabaey, et al., *Digital Integrated Circuits*, 2nd ed: Prentice Hall, 2003.

[32]   B. Davari, "CMOS technology: Present and future," presented at VLSI Circuits, 1999. Digest of Technical Papers. 1999 Symposium on, 1999.

[33]   UC Berkeley BSIM Group, (2003) "BSIM3 Manual," [Online]. Available: http://www-device.eecs.berkeley.edu/~bsim3/.

[34]   MOSIS Fabrication Service, (2003) [Online]. Available: http://www.mosis.org.

[35]   P. R. Gray, et al., *Analysis & Design of Analog Integrated Circuits*, 4th ed: John Wiley & Sons, 2001.

[36]   Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*: Cambridge University Press, 1998.

[37]   B. H. Lee, et al., "Performance enhancement on sub-70 nm strained silicon SOI MOSFETs on ultra-thin thermally mixed strained silicon/SiGe on insulator (TM-SGOI) substrate with raised S/D," *IEDM Dig. Techn. Papers*, pp. 946-948, 2002.

[38] M. J. M. Pelgrom, et al., "Matching properties of MOS transistors," *IEEE J. of Solid-State Circuits*, pp. 1433-1439, May 1989.

[39] J. Bastos, et al., "Mismatch characterization of small size MOS transistors," *Proc. Int. Conference on Microelectronic Test Structures*, pp. 271-276, 1995.

[40] K. Takeuchi, et al., "Channel engineering for the reduction of random-dopant-placement-induced threshold voltage fluctuation," *IEDM Dig. Techn. Papers*, pp. 841-844, 1997.

[41] J. A. Croon, et al., "A simple characterization method for MOS transistor matching in deep submicron technologies," *Proc. International Conference on Microelectronic Test Structures*, pp. 213-218, 2001.

[42] K. Uyttenhove and M. S. J. Steyaert, "Speed-power-accuracy tradeoff in high-speed CMOS ADCs," *IEEE Trans. Ckts. Systs II*, pp. 280-287, Apr. 2002.

[43] A. van der Ziel, "Unified presentation of 1/f noise in electron devices: fundamental 1/f noise sources," *Proceedings of the IEEE*, pp. 233-258, March 1988.

[44] A. J. Scholten, et al., "Noise modeling for RF CMOS circuit simulation," *IEEE Trans. Electron Devices*, pp. 618-632, March 2003.

[45] R. Roovers and M. S. J. Steyaert, "A 175 Ms/s, 6 b, 160 mW, 3.3 V CMOS A/D converter," *IEEE J. of Solid-State Circuits*, pp. 938-944, July 1996.

[46] K. Yoon, et al., "A 6 b 500 MSample/s CMOS flash ADC with a background interpolated auto-zeroing technique," *ISSCC Dig. Techn. Papers*, pp. 326-327, Feb. 1999.

[47] S. Tsukamoto, et al., "A CMOS 6-b, 200 MSample/s, 3 V-supply A/D converter for a PRML read channel LSI," *IEEE J. of Solid-State Circuits*, pp. 1831-1836, Nov. 1996.

[48] P. Setty, et al., "A 5.75 b 350 M sample/s or 6.75 b 150 M sample/s reconfigurable flash ADC for a PRML read channel," *ISSCC Dig. Techn. Papers*, pp. 148-149, Feb. 1998.

[49] M. Flynn and B. Sheahan, "A 400 M sample/s 6b CMOS folding and interpolating ADC," *ISSCC Dig. Techn. Papers*, pp. 150-151, Feb. 1998.

[50]   X. Jiang, et al., "A 200 MHz 6-bit folding and interpolating ADC in 0.5-µm CMOS," *Proc. ISCAS*, pp. 5-8, 1998.

[51]   S. Tsukamoto, et al., "A CMOS 6-b, 400-MSample/s ADC with error correction," *IEEE J. of Solid-State Circuits*, pp. 1939-1947, Dec. 1998.

[52]   I. Mehr and D. Dalton, "A 500-MSample/s, 6-bit Nyquist-rate ADC for disk-drive read-channel applications," *IEEE J. of Solid-State Circuits*, pp. 912-920, July 1999.

[53]   G. Geelen, "A 6 b 1.1 GSample/s CMOS A/D converter," *ISSCC Dig. Techn. Papers*, pp. 128-129, Feb. 2001.

[54]   M. Choi and A. A. Abidi, "A 6-b 1.3-Gsample/s A/D converter in 0.35µm CMOS," *IEEE J. of Solid-State Circuits*, pp. 1847-1858, Dec. 2001.

[55]   K. Uyttenhove and M. S. J. Steyaert, "A 1.8-V 6-bit 1.3-GHz flash ADC in 0.25µm CMOS," *IEEE J. of Solid-State Circuits*, pp. 1115-1122, July 2003.

[56]   C. Donovan and M. P. Flynn, "A "digital" 6-bit ADC in 0.25µm CMOS," *IEEE J. of Solid-State Circuits*, pp. 432-437, March 2002.

[57]   K. Nagaraj, et al., "A 700M Sample/s 6 b read channel A/D converter with 7 b servo mode," *ISSCC Dig. Techn. Papers*, pp. 426-427, Feb. 2000.

[58]   K. Sushihara, et al., "A 6 b 800 MSample/s CMOS A/D converter," *ISSCC Dig. Techn. Papers*, pp. 428-429, Feb. 2000.

[59]   B. Yu and W. C. Black, Jr., "A 900 MS/s 6b interleaved CMOS flash ADC," *Proc. ISCAS*, pp. 149-152, 2001.

[60]   X. Jiang, et al., "A 2Gs/s 6b ADC in 0.18µm CMOS," *ISSCC Dig. Techn. Papers*, pp. 1-10, Feb. 2003.

[61]   D. Johns and A. Hadji-Abdolhamid, "A 400-MHz 6-bit ADC with a Partial Analog Equalizer for Coaxial Cable Channels," *Proc. ESSCIRC*, 2003.

[62]   P. C. S. Scholtens and M. Vertregt, "A 6-b 1.6-Gsample/s flash ADC in 0.18-/spl mu/m CMOS using averaging termination," *IEEE J. of Solid-State Circuits*, pp. 1599-1609, Dec. 2002.

[63]   K. Nakamura, et al., "An 85 mW, 10 b, 40 Msample/s CMOS parallel-pipelined ADC," *IEEE J. of Solid-State Circuits*, pp. 173-183, March 1995.

[64]   T. B. Cho, *Low-Power Low-Voltage Analog-to-Digital Conversion Techniques Using Pipelined Architectures*: PhD Thesis, University of California, Berkeley, 1995.

[65]   B. E. Boser, *EE240 Course Notes, University of California, Berkeley*, 2003.

[66]   B.-L. Jeon, et al., "A 10 b 58 MHz CMOS A/D converter for high-speed video applications," *Proc. ASP-Design Automation Conference*, pp. 29-32, 1999.

[67]   A. Wada, et al., "A 10 b 20-Msample/s 28 mW CMOS ADC in ASIC process," *IEEE International ASIC Conference*, pp. 57-61, 1998.

[68]   A. M. Abo and P. R. Gray, "A 1.5 V, 10-bit, 14 MS/s CMOS Pipeline Analog-to-Digital Converter," *VLSI Symposium, Dig. Techn. Papers.*, pp. 166-169, 1998.

[69]   G. Chien, *High-Speed, Low-Power, Low Voltage Pipelined Analog-to-Digital Converter*: MS Thesis, University of California, Berkeley, 1996.

[70]   L. Sumanen, et al., "A 10-bit 200-MS/s CMOS parallel pipeline A/D converter," *IEEE J. of Solid-State Circuits*, pp. 1048-1055, July 2001.

[71]   I. Mehr and L. Singer, "A 55-mW, 10-bit, 40-Msample/s Nyquist-rate CMOS ADC," *IEEE J. of Solid-State Circuits*, pp. 318-325, March 2000.

[72]   D. G. Nairn, "A 10-bit, 3 V, 100 MS/s pipelined ADC," *Proc. IEEE Custom Integrated Circuits Conference (CICC)*, pp. 257-260, 2000.

[73]   J. H. Hall and D. G. Nairn, "A 100 mW 10 bit 100 MS/s all CMOS ADC," *Proc. Third Int. Conference of Advanced A/D and D/A Conversion Techniques*, pp. 5-8, 1999.

[74]   J.-S. Chiang and M.-D. Chiang, "The design of a 1.5 V, 10-bit, 10 M samples/s low power pipelined analog-to-digital converter," *Proc. ISCAS*, pp. 443-446, 2000.

[75]   D. Miyazaki, et al., "A 16 mW 30 MSample/s 10 b pipelined A/D converter using a pseudo-differential architecture," *ISSCC Dig. Techn. Papers*, pp. 174-175, Feb. 2002.

[76]   H. C. Choi, et al., "A 1.4 V 10-bit 20 MSPS pipelined A/D converter," *Proc. ISCAS*, pp. 439-442, 2000.

[77]   H. C. Choi, et al., "A 1.5 V 10-bit 25 MSPS pipelined A/D converter," *Proc. Pacific Conference on ASICs*, pp. 170-173, 1999.

[78]   Y.-I. Park, et al., "A low power 10 bit, 80 MS/s CMOS pipelined ADC at 1.8 V power supply," *Proc. ISCAS*, pp. 580-583, 2001.

[79]   B.-M. Min, et al., "A 69mW 10b 80MS/s pipelined CMOS ADC," *ISSCC Dig. Techn. Papers*, pp. 1-8, Feb. 2003.

[80]   S.-M. Yoo, et al., "A 10b 150MS/s 123mW 0.18µm CMOS pipelined ADC," *ISSCC Dig. Techn. Papers*, pp. 1-10, 2003.

[81]   A. S. Blum, et al., "A 1.2 V 10-b 100-MSamples/s A/D converter in 0.12µm CMOS," *VLSI Symposium, Dig. Tech. Papers*, pp. 326-327, 2002.

[82]   W. Yang, et al., "A 3-V 340-mW 14-b 75-Msample/s CMOS ADC with 85-dB SFDR at Nyquist input," *IEEE J. of Solid-State Circuits*, pp. 1931-6, Dec. 2001.

[83]   D. W. Cline, *Noise, Speed, and Power Trade-offs in Pipelined Analog to Digital Converters*: PhD Thesis, University of California, Berkeley, 1995.

[84]   T. B. Cho and P. R. Gray, "A 10 b, 20 Msample/s, 35 mW pipeline A/D converter," *IEEE J. of Solid-State Circuits*, pp. 166-172, March 1995.

[85]   D. W. Cline and P. R. Gray, "A power optimized 13-b 5 MSamples/s pipelined analog-to-digital converter in 1.2µm CMOS," *IEEE J. Solid-State Circuits*, pp. 294-303, March 1996.

[86]   S. H. Lewis, "Optimizing the stage resolution in pipelined, multistage, analog-to-digital converters for video-rate applications," *IEEE Trans. Ckts. Syst. II*, pp. 516-523, Aug. 1992.

[87]   L. A. Singer and T. L. Brooks, "A 14-bit 10-MHz calibration-free CMOS pipelined A/D converter," *VLSI Symposium, Dig. Techn. Papers.*, pp. 94-95, 1996.

[88]   J. Goes, et al., "Systematic design for optimization of high-speed self-calibrated pipelined A/D converters," *IEEE Trans. Ckts. Syst II*, pp. 1513-1526, Dec. 1998.

[89]   K. Gulati and H.-S. Lee, "A high-swing CMOS telescopic operational amplifier," *IEEE J. of Solid-State Circuits*, pp. 2010-2019, Dec. 1998.

[90]   P. C. Yu and H.-S. Lee, "A 2.5-V, 12-b, 5-MSample/s pipelined CMOS ADC," *IEEE J. of Solid-State Circuits*, pp. 1854-1861, Dec. 1996.

[91]   K. Poulton, et al., "A 4 Gsample/s 8b ADC in 0.35μm CMOS," *ISSCC Dig. Techn. Papers*, pp. 166-167, 2002.

[92]   R. G. Meyer, *EE242 Course Notes, University of California, Berkeley*, 2001.

[93]   P. Wambacq and W. M. Sansen, *Distortion Analysis of Analog Integrated Circuits*: Kluwer Academic Publishers, 1998.

[94]   A. Boni, et al., "A 10-b 185-MS/s track-and-hold in 0.35μm CMOS," *IEEE J. of Solid-State Circuits*, pp. 195-203, Feb. 2001.

[95]   R. Harjani, "A 455-Mb/s MR preamplifier design in a 0.8μm CMOS process," *IEEE J. of Solid-State Circuits*, pp. 862-872, June 2001.

[96]   J. J. F. Rijns, "CMOS low-distortion high-frequency variable-gain amplifier," *IEEE J. of Solid-State Circuits*, pp. 1029-1034, July 1996.

[97]   Y. Sun, et al., "Large Dynamic Range High Frequency Fully Differential CMOS Transconductance Amplifier," *Analog Integrated Circuits and Signal Processing*, pp. 247-255, March 2003.

[98]   V. I. Prodanov, "V-I converters with transconductance proportional to bias current in any technology," *Proc. ISCAS*, pp. 201-204, 2000.

[99]   E. G. Soenen and R. L. Geiger, "An architecture and an algorithm for fully digital correction of monolithic pipelined ADCs," *IEEE Trans. Ckts. Syst. II*, pp. 143-153, March 1995.

[100] G. Birkhoff and S. Mac Lane, *A Survey of Modern Algebra*, 5th ed. New York: Macmillan, 1996.

[101] J. Tsimbinos, *Identification and Compensation of Nonlinear Distortion*. The Levels: University of South Australia, 1995.

[102] T.-H. Shu, et al., "A 13-b 10-Msample/s ADC digitally calibrated with oversampling delta-sigma converter," *IEEE J. of Solid-State Circuits*, pp. 443-452, April 1995.

[103] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*: Englewood Cliffs, NJ: Prentice-Hall, 1985.

[104] L. Singer, et al., "A 12 b 65 MSample/s CMOS ADC with 82 dB SFDR at 120 MHz," *ISSCC Dig. Techn. Papers*, pp. 38-9, Feb. 2000.

[105] I. E. Opris, et al., "A single-ended 12-bit 20 Msample/s self-calibrating pipeline A/D converter," *IEEE J. of Solid-State Circuits*, pp. 1898-1903, Dec. 1998.

[106] J. Altet, et al., "Thermal coupling in integrated circuits: application to thermal testing," *IEEE J. of Solid-State Circuits*, pp. 81-91, Jan. 2001.

[107] R. Gharpurey, *Modeling and Analysis of Substrate Coupling in Integrated Circuits*: PhD Thesis, University of California, Berkeley, 1995.

[108] Analog Devices, (2003) "AD9235 Data Sheet," [Online]. Available: www.analog.com.

[109] S. Limotyrakis, et al., "Analysis and simulation of distortion in folding and interpolating A/D converters," *IEEE Trans. Ckts. Syst II*, pp. 161-169, March 2002.

[110] J. Larrabee, et al., "Using sine wave histograms to estimate analog-to-digital converter dynamic error functions," *IEEE Transactions on Instrumentation and Measurement*, pp. 1448-1456, June 1998.

[111] D. M. Hummels, et al., "Discrete-time dynamic compensation of analog-to-digital converters," *Proc. ISCAS*, pp. 1144-1147, 1993.

[112] C. J. Stone, *A Course in Probability and Statistics*: Duxbury, 1991.

[113] B. C. Arnold, et al., *A first course in order statistics*. New York: Wiley, 1992.

[114] A. M. Mood, et al., *Introduction to the theory of statistics*. New York: McGraw-Hill, 1973.

This page intentionally left blank

Index