

PRECISION TEMPERATURE SENSORS IN CMOS TECHNOLOGY

ANALOG CIRCUITS AND SIGNAL PROCESSING SERIES

Consulting Editor: Mohammed Ismail. Ohio State University

Related Titles:

PRECISION TEMPERATURE SENSORS IN CMOS TECHNOLOGY

Pertjts, Michiel A.P., Huijsing, Johan H.

ISBN: 1-4020-5257-X

RF POWER AMPLIFIERS FOR MOBILE COMMUNICATIONS

Reynaert, Patrick, Steyaert, Michiel

ISBN: 1-4020-5116-6

ADAPTIVE LOW-POWER CIRCUITS FOR WIRELESS COMMUNICATIONS

Tasic, Aleksandar, Serdijn, Wouter A., Long, John R.

ISBN: 1-4020-5249-9

IQ CALIBRATION TECHNIQUES FOR CMOS RADIO TRANCEIVERS

Chen, Sao-Jie, Hsieh, Yong-Hsiang

ISBN: 1-4020-5082-8

CMOS CURRENT-MODE CIRCUITS FOR DATA COMMUNICATIONS

Yuan, Fei

ISBN: 0-387-29758-8

ADVANCED DESIGN TECHNIQUES FOR RF POWER AMPLIFIERS

Rudiakova, A.N., Krizhanovski, V.

ISBN 1-4020-4638-3

CMOS CASCADE SIGMA-DELTA MODULATORS FOR SENSORS AND TELECOM

del Río, R., Medeiro, F., Pérez-Verdú, B., de la Rosa, J.M., Rodríguez-Vázquez, A.

ISBN 1-4020-4775-4

Titles in former series International Series in Engineering and Computer Science:

SIGMA DELTA A/D CONVERSION FOR SIGNAL CONDITIONING

Philips, K., van Roermund, A.H.M.

Vol. 874, ISBN 1-4020-4679-0

CALIBRATION TECHNIQUES IN NYQUIST A/D CONVERTERS

van der Ploeg, H., Nauta, B.

Vol. 873, ISBN 1-4020-4634-0

ADAPTIVE TECHNIQUES FOR MIXED SIGNAL SYSTEM ON CHIP

Fayed, A., Ismail, M.

Vol. 872, ISBN 0-387-32154-3

WIDE-BANDWIDTH HIGH-DYNAMIC RANGE D/A CONVERTERS

Doris, Konstantinos, van Roermund, Arthur, Leenaerts, Domine

Vol. 871, ISBN: 0-387-30415-0

METHODOLOGY FOR THE DIGITAL CALIBRATION OF ANALOG CIRCUITS AND SYSTEMS: WITH CASE STUDIES

Pastre, Marc, Kayal, Maher

Vol. 870, ISBN: 1-4020-4252-3

HIGH-SPEED PHOTODIODES IN STANDARD CMOS TECHNOLOGY

Radovanovic, Sasa, Annema, Anne-Johan, Nauta, Bram

Vol. 869, ISBN: 0-387-28591-1

LOW-POWER LOW-VOLTAGE SIGMA-DELTA MODULATORS IN NANOMETER CMOS

Yao, Libin, Steyaert, Michiel, Sansen, Willy

Vol. 868, ISBN: 1-4020-4139-X

DESIGN OF VERY HIGH-FREQUENCY MULTIRATE SWITCHED-CAPACITOR CIRCUITS

U, Seng Pan, Martins, Rui Paulo, Epifânio da Franca, José

Vol. 867, ISBN: 0-387-26121-4

DYNAMIC CHARACTERISATION OF ANALOGUE-TO-DIGITAL CONVERTERS

Dallet, Dominique; Machado da Silva, José (Eds.)

Vol. 860, ISBN: 0-387-25902-3

ANALOG DESIGN ESSENTIALS

Sansen, Willy

Vol. 859, ISBN: 0-387-25746-2

DESIGN OF WIRELESS AUTONOMOUS DATALOGGER IC'S

Claes and Sansen

Vol. 854, ISBN: 1-4020-3208-0

MATCHING PROPERTIES OF DEEP SUB-MICRON MOS TRANSISTORS

Croon, Sansen, Maes

Vol. 851, ISBN: 0-387-24314-3

LNA-ESD CO-DESIGN FOR FULLY INTEGRATED CMOS WIRELESS RECEIVERS

Leroux and Steyaert

Vol. 843, ISBN: 1-4020-3190-4

SYSTEMATIC MODELING AND ANALYSIS OF TELECOM FRONTENDS AND THEIR BUILDING BLOCKS

Vanassche, Gielen, Sansen

Vol. 842, ISBN: 1-4020-3173-4

PRECISION TEMPERATURE SENSORS IN CMOS TECHNOLOGY

by

Michiel A.P. Pertijs

National Semiconductor, Delft, The Netherlands

and

Johan H. Huijsing

Delft University of Technology, The Netherlands



A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-5257-X (HB)

ISBN-13 978-1-4020-5257-6 (HB)

ISBN-10 1-4020-5258-8 (e-book)

ISBN-13 978-1-4020-5258-3 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

All Rights Reserved

© 2006 Springer

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Contents

Acknowledgment	xi
1. INTRODUCTION	1
1.1 Motivation and Objectives	1
1.2 Basic Principles	3
1.3 Context of the Research	5
1.4 Challenges	6
1.5 Organization of the Book	7
References	8
2. CHARACTERISTICS OF BIPOLAR TRANSISTORS	11
2.1 Introduction	11
2.1.1 The Ideal Diode Characteristic	11
2.1.2 Non-Idealities of Diodes	13
2.2 Bipolar Transistor Physics	14
2.2.1 Sign Conventions	14
2.2.2 The Ideal $I_C - V_{BE}$ Characteristic	15
2.2.3 Non-Idealities of the $I_C - V_{BE}$ Characteristic	16
2.2.4 Non-Idealities of the $I_E - V_{BE}$ Characteristic	18
2.3 Temperature Characteristics of Bipolar Transistors	20
2.3.1 Temperature Dependency of the Saturation Current	21
2.3.2 Temperature Dependency of the Current Gain	23
2.4 Bipolar Transistors in Standard CMOS Technology	24
2.4.1 Lateral pnp Transistors	24
2.4.2 Substrate pnp Transistors	26
2.5 Processing Spread	28
2.5.1 Spread of the Saturation Current	29

2.5.2	Spread of the Current Gain	31
2.6	Sensitivity to Mechanical Stress	32
2.6.1	Causes of Mechanical Stress	33
2.6.2	Stress-Induced Changes in the Saturation Current	34
2.6.3	Stress-Induced Changes in the Current Gain	35
2.7	Effect of Series Resistances and Base-Width Modulation	36
2.7.1	Series Resistances	36
2.7.2	Forward Early Effect	37
2.7.3	Reverse Early Effect	38
2.8	Effect of Variations in the Bias Current	40
2.8.1	Resistors in Standard CMOS Technology	40
2.8.2	Temperature Dependency of the Bias Resistor	41
2.8.3	Spread of the Bias Resistor	43
2.8.4	Stress-Induced Changes in the Bias Resistor	45
2.9	Conclusions	46
	References	48
3.	RATIOMETRIC TEMPERATURE MEASUREMENT USING BIPOLAR TRANSISTORS	51
3.1	Introduction	51
3.1.1	Combining V_{BE} and ΔV_{BE}	51
3.1.2	Error Budgeting	54
3.1.3	Errors in V_{BE} , ΔV_{BE} and α	55
3.1.4	PTAT Errors in V_{BE}	56
3.2	Generating an Accurate Current-Density Ratio	58
3.2.1	Errors due to Mismatch	58
3.2.2	Dynamic Element Matching	59
3.2.3	Errors due to Finite Output Impedance	62
3.3	Generating an Accurate Bias Current	63
3.3.1	Structure of Bias Circuits	63
3.3.2	PTAT/R Bias Circuit	66
3.3.3	Compensation for Processing Spread	67
3.4	Trimming	69
3.4.1	Calculating Trimming Parameters	69
3.4.2	Trimming Circuits	70
3.4.3	Trimming after Packaging	74
3.4.4	Non-Volatile Memory Technology	76
3.5	Curvature Correction	78

3.5.1	Errors due to Curvature	78
3.5.2	Comparison to Voltage References	79
3.5.3	Curvature-Correction Techniques for Bandgap Voltage References	80
3.5.4	Ratiometric Curvature Correction	88
3.5.5	Higher-Order Ratiometric Curvature Correction	89
3.5.6	Other Curvature-Correction Techniques	91
3.6	Compensation for Finite Current-Gain	92
3.6.1	Errors due to Finite Current-Gain	92
3.6.2	Current-Gain-Dependent Biasing	92
3.7	Series-Resistance Compensation	95
3.7.1	Errors due to Series Resistances	95
3.7.2	Instantaneous Compensation	96
3.7.3	Sequential Compensation	98
3.8	Conclusions	99
	References	102
4.	SIGMA-DELTA ANALOG-TO-DIGITAL CONVERSION	107
4.1	Introduction	107
4.1.1	Requirements	107
4.1.2	Direct versus Indirect Conversion	110
4.1.3	Charge Balancing	111
4.1.4	Synchronous versus Asynchronous Modulation	112
4.2	Operating Principles of Sigma-Delta ADCs	115
4.2.1	Sampling and Quantization	115
4.2.2	Oversampling	116
4.2.3	Noise Shaping	117
4.2.4	Linear Model	118
4.2.5	Incremental Operation	120
4.3	First-Order Sigma-Delta Modulators	121
4.3.1	Topology	121
4.3.2	Noise Shaping	121
4.3.3	Resolution	121
4.3.4	Leakage	125
4.3.5	Initialization	128
4.4	Second-Order Sigma-Delta Modulators	129
4.4.1	Cascading versus Higher-Order Loop Filters	129
4.4.2	Topology	131
4.4.3	Stability	132

4.4.4	Noise Shaping	134
4.4.5	Resolution	134
4.4.6	Leakage	136
4.5	Decimation Filters	137
4.5.1	Filters Matched to the Loop Filter	138
4.5.2	Filters Based on Window Functions	139
4.5.3	Linear Scaling of the Conversion Result	142
4.5.4	Compensating for Non-Linearity	143
4.6	Filtering of Dynamic Error Signals	146
4.6.1	Normal-Mode Rejection	146
4.6.2	Bitstream-Controlled Timing of Dynamic Error Signals	149
4.7	Conclusions	153
	References	155
5.	PRECISION CIRCUIT TECHNIQUES	159
5.1	Introduction	159
5.1.1	Methods of Voltage-to-Charge Conversion	159
5.1.2	Maximum Power Consumption Based on Self-Heating	160
5.1.3	Per-Cycle Analysis of Noise	162
5.1.4	Noise of the Bipolar Front-End	163
5.2	Continuous-Time Circuitry	165
5.2.1	Implementation of Charge Balancing	165
5.2.2	Accuracy	167
5.2.3	Noise	172
5.2.4	Power Consumption	174
5.2.5	Chopping	175
5.2.6	Dynamic Element Matching	180
5.3	Switched-Capacitor Circuitry	182
5.3.1	Implementation of Charge Balancing	182
5.3.2	Accuracy	184
5.3.3	Noise	187
5.3.4	Power Consumption	190
5.3.5	Autozeroing	191
5.3.6	Dynamic Element Matching	194
5.4	Advanced Offset Cancellation Techniques	196
5.4.1	Charge-Injection Compensation	196
5.4.2	Advanced Chopping Techniques	199
5.4.3	Advanced Autozeroing Techniques	201

5.4.4	System-Level Techniques	204
5.5	Conclusions	207
	References	209
6.	CALIBRATION TECHNIQUES	213
6.1	Introduction	213
6.1.1	Definition of Calibration	213
6.1.2	Extrapolation from Calibration Points	215
6.2	Conventional Calibration Techniques	217
6.3	Batch Calibration	217
6.4	Calibration based on ΔV_{BE} Measurement	218
6.4.1	Principle	218
6.4.2	Implementation	219
6.4.3	Accuracy	220
6.5	Voltage Reference Calibration	221
6.5.1	Principle	221
6.5.2	Implementation	222
6.5.3	Accuracy	223
6.6	Conclusions	223
	References	224
7.	REALIZATIONS	227
7.1	A Batch-Calibrated CMOS Smart Temperature Sensor	227
7.1.1	Overview	227
7.1.2	Charge-Balancing Operation	229
7.1.3	Curvature Correction	231
7.1.4	Sinking V-I Converter for ΔV_{BE}	232
7.1.5	Sourcing V-I Converter for V_{BE}	234
7.1.6	Experimental Results	235
7.2	A CMOS Smart Temperature Sensor with a 3σ Inaccuracy of $\pm 0.5^\circ\text{C}$ from -50°C to 120°C	236
7.2.1	Overview	236
7.2.2	Sigma-Delta Modulator	238
7.2.3	Experimental Results	240
7.3	A CMOS Smart Temperature Sensor with a 3σ Inaccuracy of $\pm 0.1^\circ\text{C}$ from -55°C to 125°C	244
7.3.1	Overview	244
7.3.2	Charge-Balancing Operation	246
7.3.3	Dynamic Element Matching	248

7.3.4	Modulated Bias Current Trimming	249
7.3.5	Precision Bias Circuit	250
7.3.6	Sigma-Delta Modulator	252
7.3.7	Timing and Decimation Filter	257
7.3.8	Calibration	258
7.3.9	Experimental Results	260
7.4	Benchmark	268
	References	268
8.	CONCLUSIONS	271
8.1	Main Findings	271
8.2	Other Applications of this Work	272
8.3	Future Work	272
	References	274
	Appendices	275
A	Derivation of Mismatch-Related Errors	275
A.1	Errors in ΔV_{BE}	275
A.1.1	Without DEM	275
A.1.2	With DEM	276
B	Resolution Limits of Sigma-Delta Modulators with a DC Input	277
B.1	First-Order Modulator	277
B.1.1	Time-Domain Description	277
B.1.2	Resolution Limit without Leakage	277
B.1.3	Resolution Limit with Leakage	278
B.2	Second-Order Single-Loop Modulator	279
B.2.1	Time-Domain Description	279
B.2.2	Resolution Limit without Leakage	279
B.2.3	Resolution Limit with Leakage	281
	References	281
C	Non-Exponential Settling Transients	283
C.1	Problem Description	283
C.2	Settling Transients from $V_{BE1} \neq 0$ to V_{BE2}	284
C.3	Settling Transients from $V_{BE1} = 0$ to V_{BE2}	286
	Summary	291
	About the Authors	295
	Index	297

ACKNOWLEDGMENT

This book started life as a Ph.D. thesis written at the Electronic Instrumentation Laboratory of Delft University of Technology, where I spent an exiting, productive and very enjoyable period of about five years. I would like to thank the people who have made these years so enjoyable.

I would like to start by thanking Han Huijsing, who was my advisor during my Ph.D. project. I am grateful for his support, encouragement, and his trust in me. I also thank Gerard Meijer, for his support and useful feedback on my work. I was very fortunate to have the ‘godfather’ of integrated temperature sensing so close by.

I thank the members of the Electronic Instrumentation Lab for creating a great working environment. My special thanks go to my roommates and fellow crocs Kofi and Martijn. I greatly enjoyed working with you, and have fond memories of all the technical and non-technical creative moments we shared. I also want to thank my other colleagues of the Huijsing/Makinwa group, Frerik, Paulo, Svetla, Ovidiu, Jeroen, David, and André, for their support and friendship. Special thanks also go to Maureen, Harry, Piet, Jeroen, Ger, Jeff, and Antoon for their indispensable technical support. My thanks also go to Inge, Trudie, Evelyn and Willem, whose administrative support keeps the lab going.

I also thank Anton Bakker, who was my mentor during my M.Sc. project, and provided support for my Ph.D. project from within Philips. His Ph.D. work on CMOS temperature sensors formed the basis for my work, and many of the results presented in this book would not have been possible without him. I very much enjoyed the times we spent together in the US, and the many games of squash we played together with Gian, Marto-Jan and Kofi.

I thank Greta Milczanowska and Nico Beylemans of IMEC, who carefully prepared my designs for processing through the Europractice IC service. Thanks also go to Wim van der Vlist en Ruud Klerks for bonding and packaging many of my chips.

I am grateful to the (former) employees of the Philips Standard Analog Business Line in Tempe, Arizona, for making my visits there so enjoyable. In particular, I would like to mention Andrea Niederkorn, Bill McKillip, and Jason Ma, who designed parts of the chip described in Section 7.2. Thanks also go to Hung Nguyen, who performed many measurements, and introduced me to the practical problems of temperature sensor testing. Thanks also to Don Remsen and Kevin Thiele for making my practical training at Philips Sunnyvale possible.

I would like to thank the Dutch Technology Foundation STW for their financial support. I thank Rolf the Boer of Smartec for his support and the regular discussions we had. Thanks also go to Jeff West of the Philips Interface Products Business Line for his continued interest in my work.

I am very grateful to the people who proofread (parts of) the bulky manuscript of this book: Kofi, Gerard, Martijn, Frerik, André, and Mirjam Nieman. Your comments have undoubtedly improved the readability of the text.

I thank my friends, in particular Martijn en Dubi, for sharing the ups and downs of Ph.D. student life, and for providing many opportunities to get away from it all. I am very grateful to my parents, who have always been there for me, and have supported and encouraged me. Finally, I thank Hannah for her love, support and understanding.

Michiel Pertijs
Delft, May 2006

Chapter 1

INTRODUCTION

The low cost and direct digital output of CMOS smart temperature sensors are important advantages compared to conventional temperature sensors. This book addresses the main problem that nevertheless prevents widespread application of CMOS smart temperature sensors: their relatively poor absolute accuracy. Several new techniques are introduced to improve this accuracy. The effectiveness of these techniques is demonstrated using three prototypes. The final prototype achieves an inaccuracy of $\pm 0.1^\circ\text{C}$ over the military temperature range, which is a significant improvement in the state of the art. Since smart temperature sensors have been the subject of academic and industrial research for more than two decades, an overview of existing knowledge and techniques is also provided throughout the book.

In this introductory chapter, the motivation and objectives of this work are described. This is followed by a review of the basic operating principles of CMOS smart temperature sensors, and a brief overview of previous work. The challenges are then described that need to be met in order to improve the accuracy of CMOS smart temperature sensors while maintaining their cost advantage. Finally, the structure of the rest of the book is introduced.

1.1 Motivation and Objectives

Temperature sensors are widely applied in measurement, instrumentation and control systems. In an average household, at least a dozen temperature sensors can be found in various places, ranging from the coffee machine, to the heating system, to the car. Given this large market, it makes sense to try to fabricate temperature sensors in integrated circuit (IC) technology, as this is ideally suited for the volume production of low-cost products. Moreover, a temperature sensor fabricated in IC technology can be combined with interface electronics on a single chip. Such ‘smart’ sensors have distinct

advantages compared to conventional sensors: they can directly communicate with a microcomputer in a standardized digital format, thus reducing the complexity and increasing the modularity of the system in which they are applied. In addition, the local processing of the sensor signal (amplification, analog-to-digital conversion) makes the measurement more robust to interference [1, 2].

In spite of these advantages, only a minority of the temperature sensors applied today are smart sensors. The semiconductor industry only became successful in marketing smart temperature sensors when an application ‘close to home’ presented itself: thermal management in personal computers and laptops [3]. Because of the steady increase in heat dissipation that has accompanied the increasing processing power of microprocessors, temperature sensors are needed to track the processor’s temperature and regulate its cooling fan. During the last decade, this application has given an enormous boost to the development of smart temperature sensors.

The relatively limited use of smart temperature sensors in other applications can be partially attributed to the success of conventional sensors: platinum resistors, thermistors and thermopiles have been used successfully for decades, so designers are hesitant to adopt a new, ‘unproven’ technology, in spite of its possible advantages. Also, for some applications, the operating range of integrated temperature sensors (typically -55°C to 125°C) is too restricted. For many other applications, however, the limited accuracy of smart temperature sensors is the most important obstacle.

There are two main reasons for the limited accuracy of smart temperature sensors. To keep production costs low, smart temperature sensors are often produced in standard CMOS technology, which has been developed for mainstream digital products, not precision analog products. In addition, their temperature error is typically measured (calibrated) and corrected at not more than one temperature. Over their full operating range, the resulting inaccuracy is then typically not better than $\pm 2.0^{\circ}\text{C}$ [1]. For comparison, the inaccuracy that can be obtained with a class-A platinum resistor over that range is $\pm 0.5^{\circ}\text{C}$ [4]. Until recently, such inaccuracy could only be obtained from smart temperature sensors by calibrating them at several temperatures, which undoes much of their cost advantage.

As the use of standard CMOS technology and calibration at not more than one temperature are required to keep production costs low, improvements in the accuracy of smart temperature sensors should be sought in improvements in sensor design and in the calibration procedure. This book presents such improvements. Through a combination of existing precision interfacing techniques and several new techniques, a significant improvement in accuracy has been obtained. This has been demonstrated by the realization of several prototypes, which will be described at the end of the book. The most advanced

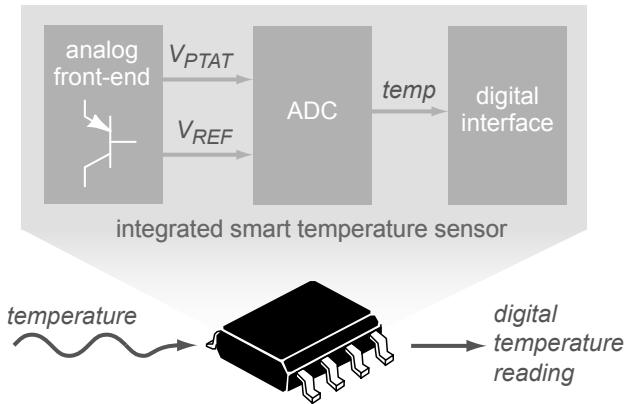


Figure 1.1. Block diagram of an integrated smart temperature sensor.

prototype has an inaccuracy of $\pm 0.1^\circ\text{C}$ over the temperature range of -55°C to 125°C [5]. This is, to date, the best reported accuracy for this type of sensors.

With such improved designs, smart temperature sensors can be produced that can compete with conventional temperature sensors in terms of both cost and accuracy. Thus, the mentioned advantages of smart sensors will become available to a wider range of applications.

1.2 Basic Principles

As virtually every device characteristic in an integrated circuit is temperature dependent, there are numerous ways of making integrated temperature sensors. Sensors have been reported that are based on the temperature dependency of resistors, MOS transistors [6], thermal delay-lines [7], tunneling diodes [8] and, predominantly, bipolar transistors [9]. The output of such sensors is typically an analog signal: a temperature-dependent voltage, current, period or frequency.

A single temperature-dependent analog signal alone, however, is not enough to realize a *smart* temperature sensor. This is because the output of a smart temperature sensor should be a digital representation of its temperature. To produce such a representation, a temperature-dependent signal needs to be compared with a reference signal, that is, a *ratiometric* measurement is needed.

Most smart temperature sensors make use of the characteristics of bipolar transistors. These characteristics are based on two voltages that can play the role of the signals required for ratiometric temperature measurement: the thermal voltage kT/q (where k is Boltzmann's constant, T is the absolute temperature, and q is the charge of an electron) and the silicon bandgap voltage V_{g0} . The thermal voltage can be used to generate a voltage V_{PTAT} that is proportional

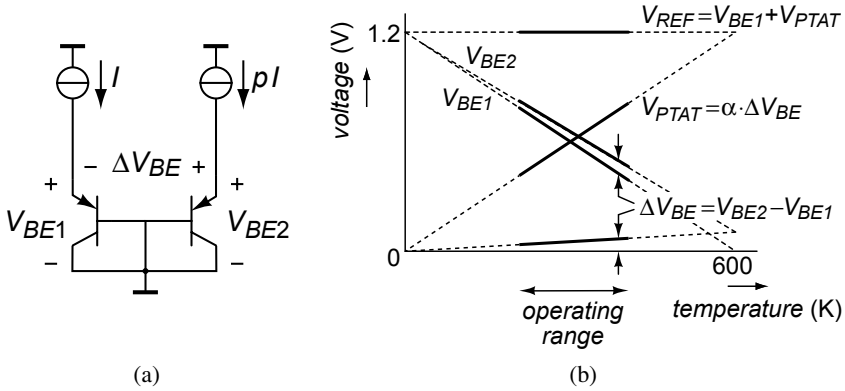


Figure 1.2. Operating principle of a CMOS smart temperature sensor: (a) two diode-connected pnp transistors are biased at a well-defined current density ratio $1 : p$; (b) their base-emitter voltages are used to generate a voltage proportional to absolute temperature V_{PTAT} and a bandgap reference V_{REF} , the ratio of which is a measure of temperature.

to absolute temperature (PTAT), while the bandgap voltage is the basis for generating a temperature-independent reference voltage V_{REF} [9].

In an integrated smart temperature sensor (see Figure 1.1), a number of bipolar transistors are combined with precision interface circuitry in an analog front-end to extract these voltages. A digital representation of their ratio is then determined by an analog-to-digital converter (ADC). This ratio is a measure of the chip's temperature and is communicated to the outside world (e.g. to a microcomputer) by means of a digital interface. This interface can be, for instance, a standardized serial interface, such as I²C, that allows the sensor to communicate over a small number of wires, possibly shared by multiple sensors [2].

Figure 1.2 illustrates the operating principle of the analog front-end. In CMOS technology, the bipolar transistors are usually substrate pnp transistors, which are parasitic devices that are present as a side-effect of the MOS transistors for which the technology was designed. The PTAT voltage is generated from the difference in base-emitter voltage ΔV_{BE} between two such transistors biased at different current densities (Figure 1.2a). If the ratio p of the bias currents is well-defined, this difference is accurately PTAT. It is, however, quite small ($0.1 - 0.25 \text{ mV} / \text{K}$) and therefore needs to be amplified to get a useful voltage V_{PTAT} .

The reference voltage is based on the absolute base-emitter voltage of a bipolar transistor, rather than on a difference. Extrapolated to 0 K, this base-emitter voltage is equal to the silicon bandgap voltage of about 1.2 V (Figure 1.2b). From there, it decreases by about $2 \text{ mV} / \text{K}$. To compensate for this decrease, a voltage $\alpha \cdot \Delta V_{BE}$ is added to it, resulting in a voltage V_{REF} that is

essentially temperature-independent [9]. Since V_{REF} is nominally equal to the silicon bandgap voltage, such a reference is referred to as a ‘bandgap reference’.

1.3 Context of the Research

General-purpose bandgap references were invented before temperature sensors based on the same principles came around. The first bandgap reference was introduced by Hilbiber in 1964 [10] and made practical in the seminal work of Widlar in 1971 [11]. The first analog sensor that used ΔV_{BE} as a measure of temperature was presented by Verster in 1968 [12], while the first integrated version of such a sensor was described by Dobkin [13]. A lot of pioneering work was done by Meijer [14]. The first *smart* temperature sensors (based on a ratiometric measurement and with a digital interface) were introduced around 1985 and targeted the relatively small temperature range required to measure human body temperature [15, 16]. They were fabricated in bipolar technology, which dominated the industry at the time. A more general-purpose smart sensor was introduced by Meijer in 1989 [17].

During the 1990’s, CMOS technology became mainstream. Due to the ever-increasing demand for digital circuits, CMOS technology became available at much lower costs than bipolar technology. Though digital CMOS technology is typically not designed to include bipolar transistors, such transistors usually are available as parasitic devices. These devices come in two flavors, lateral and vertical, both of which have, for general purposes, a poorer performance than their counterparts in bipolar technology.

Already in 1983, Vittoz investigated the characteristics and application of lateral bipolar transistors and pointed out that they could be used in CMOS temperature sensors [18]. The first CMOS smart temperature sensor based on lateral bipolar transistors was published by Krummenacher in 1990 [19].

The temperature dependency of vertical bipolar transistors has been studied by Wang, who has shown that they are very suitable for making bandgap references and temperature sensors [20]. Moreover, Fruett [21] and Creemer [22] have shown in their study of the piezo-junction effect that vertical pnp transistors in CMOS are relatively insensitive to mechanical stress. This makes them the device of choice, especially for temperature sensors that are exposed to stress due to low-cost plastic packaging.

One of the first CMOS smart temperature sensors based on vertical bipolar transistors was presented by Bakker in 1996 [23]. His work focused on solving the problem that amplifiers in CMOS technology have a much larger offset than their counterparts in bipolar technology, which prevents accurate amplification of ΔV_{BE} . With his nested-chopper technique, he reduced the offset to the 100 nV level, which reduces the associated temperature errors to negligible levels [1]. The dynamic element matching technique, introduced by Klaassen [24] and Van der Plassche [25], can be employed to make amplifiers with an

accurate gain. Its applicability to temperature sensors has been demonstrated by Meijer *et al.* [26].

Most recent work on smart temperature sensors has been performed in industry, fueled by their above-mentioned application in PCs and laptops. The biggest players in this field are Analog Devices [27], National Semiconductor [28] and Maxim [29], who have all developed an extensive line of thermal management chips. For this application, an accuracy of $\pm 1.0^\circ\text{C}$ is usually sufficient, although a higher accuracy for the same price is of course a competitive advantage. Developments have mainly focused, however, on incorporating extra functionality on the sensor chip, such as voltage monitoring and temperature measurement of a remote diode located on the microprocessor chip [3].

1.4 Challenges

The accuracy that can be reached with a smart temperature sensor is ultimately determined by the accuracy of the temperature characteristics of the bipolar transistors. How much trimming is needed depends on the reproducibility of these characteristics in production. The difference in base-emitter voltage ΔV_{BE} is very reproducible, because the components that depend on production parameters cancel when two base-emitter voltages are subtracted. As a result, as will be shown in this book, temperature errors as a result of inaccuracy of ΔV_{BE} can be made less than $\pm 0.1^\circ\text{C}$ when transistor non-idealities are taken into account [30].

A single base-emitter voltage V_{BE} (and therefore also the reference voltage V_{REF} that is derived from it), in contrast, is much less reproducible. This is a result of production tolerances in the CMOS production process, and cannot be avoided by clever circuit design. It results in temperature errors in the order of at least $\pm 1^\circ\text{C}$. Fortunately, because there is essentially only one degree of freedom in the production spread of V_{BE} , a calibration at a single temperature provides sufficient information to correct the resulting error over the full operating range [26].

There are therefore two main challenges in the design of a high-accuracy CMOS smart temperature sensor. First, the sensor's electronics have to be designed in such a way that the spread of V_{BE} is the only significant error source. Second, a cost-effective calibration technique has to be developed that can be used to determine the temperature error resulting from the spread of V_{BE} .

The first challenge is one of precision analog circuit design, and can be broken down in two sub-challenges (see Figure 1.3). The first is the design of a circuit that generates well-defined bias currents for the bipolar transistors and operates them in a region most suited for accurate temperature sensing. This is the main topic of Chapter 3 of this book. The second is the design of an ADC that accurately processes the base-emitter voltages and generates the desired

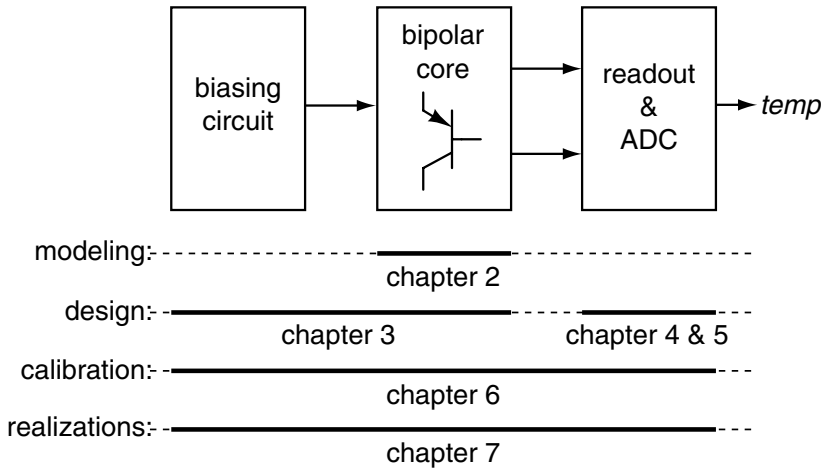


Figure 1.3. Sub-blocks of a smart temperature sensor, with an indication of where they are discussed in the book.

digital temperature reading. The system-level design of such an ADC will be discussed in Chapter 4, while the circuit-level design is described in Chapter 5.

The second challenge, the development of a cost-effective technique, is related to the design of the production line. Calibration of smart temperature sensors can either take place on wafer-level, i.e. before the sensors are diced and packaged, or after packaging. Calibration at wafer-level is typically cheaper, because a large number of sensors can be calibrated at the same time. The subsequent packaging, however, introduces additional errors due to mechanical stress, especially if a low-cost plastic package is used [31].

It will be shown that for inaccuracies below $\pm 0.5^\circ\text{C}$, it is required to calibrate the sensors individually after packaging. If this is done in a traditional way, by comparing a sensor's output with an accurate thermometer, a thermal stabilization time in the order of minutes is needed to guarantee a small enough temperature difference between the sensor and the thermometer. This is very long compared to the time typically spent on the electrical production tests performed after packaging, which is in the order of seconds. To keep production costs low, an alternative technique is needed that can be completed in the same time frame as the electrical tests. Such techniques are introduced in Chapter 6.

1.5 Organization of the Book

The organization of this book is as follows (see Figure 1.3). In order to improve the design of CMOS smart temperature sensors, a solid understanding of the characteristics of the sensor element, i.e. the bipolar core, is needed.

This is developed in Chapter 2, by reviewing the relevant physical backgrounds of bipolar transistors, and by modeling the non-idealities of V_{BE} and ΔV_{BE} . This establishes a lower bound on the overall accuracy that can be obtained. Trimming is shown to be indispensable for inaccuracy below $\pm 1.0^\circ\text{C}$, but it is also shown that a calibration at a single temperature can be sufficient to achieve a high accuracy over a wide temperature range.

Given the models for the behavior of the bipolar core, design techniques for the other parts of the sensor are developed in Chapters 3, 4 and 5. In Chapter 3, the combination of the biasing circuit and the bipolar core is considered. This ‘front-end’ of the sensor generates the two voltages V_{BE} and ΔV_{BE} . The techniques discussed in Chapter 3 focus on reducing or compensating for the non-idealities of these voltages.

Chapter 4 discusses the system-level design of the ADC that processes V_{BE} and ΔV_{BE} . Sigma-delta ADCs are found to be best suited for precision smart temperature sensors. The principles of first- and second-order sigma-delta converters are described. Details of the design of a decimation filter that can directly provide an output in $^\circ\text{C}$ are also included. Finally, the filtering properties of sigma-delta ADC are discussed, which are important if dynamic error correction techniques are applied in the front-end of the sensor, which generate modulated signals at the input of the ADC that need to be filtered out.

Sigma-delta ADCs can be realized using continuous-time or switched-capacitor circuits. In Chapter 5, these alternatives are analyzed and compared in terms of accuracy and noise. Offset cancellation and dynamic element matching techniques are discussed, which are needed for accurate conversion in the presence of mismatch.

In Chapter 6, three low-cost calibration techniques are described and compared with conventional techniques.

In Chapter 7, the techniques discussed in Chapter 3-6 are combined in three realizations, which achieve inaccuracies of $\pm 1.5^\circ\text{C}$, $\pm 0.5^\circ\text{C}$, and finally $\pm 0.1^\circ\text{C}$. Details of these realizations are described and experimental results are presented. The results are compared with previous work, showing that the realized sensors have the highest reported accuracy to date.

The book ends with conclusions and a summary. Special sections have been devoted to other applications of the work and future work on smart temperature sensors.

References

- [1] A. Bakker and J. H. Huijsing, *High-Accuracy CMOS Smart Temperature Sensors*. Boston: Kluwer Academic Publishers, 2000.
- [2] J. H. Huijsing, F. R. Riedijk, and G. van der Horn, “Developments in integrated smart sensors,” *Sensors and Actuators*, vol. 43, no. 1-3, pp. 276–288, May 1994.

- [3] J. Steele, "ACPI thermal sensing and control in the PC," in *Proc. WESCON*, 1998, pp. 169–182.
- [4] G. C. M. Meijer and A. W. van Herwaarden, *Thermal Sensors*. Bristol, UK: IOP Publishing, 1994.
- [5] M. A. P. Pertijs, K. A. A. Makinwa, and J. H. Huijsing, "A CMOS temperature sensor with a 3σ inaccuracy of $\pm 0.1^\circ\text{C}$ from -55°C to 125°C ," in *Dig. Techn. Papers ISSCC*, Feb. 2005, pp. 238–239, 596.
- [6] I. M. Filanovsky and W. Lee, "Two temperature sensors with signal-conditioning amplifiers realized in BiCMOS technology," *Sensors and Actuators*, vol. 77, pp. 45–53, Sept. 1999.
- [7] V. Székely and M. Rencz, "A new monolithic temperature sensor: The thermal feedback oscillator," in *Proc. Transducers*, 1995, pp. 124–127.
- [8] Y. Shih, S. Lin, T. Wang, and J. Hwu, "High sensitive and wide detecting range MOS tunneling temperature sensors for on-chip temperature detection," *IEEE Transactions on Electron Devices*, vol. 51, no. 9, pp. 1514–1521, Sept. 2004.
- [9] G. C. M. Meijer, "Thermal sensors based on transistors," *Sensors and Actuators*, vol. 10, pp. 103–125, Sept. 1986.
- [10] D. F. Hilbiber, "A new semiconductor voltage standard," in *Dig. Techn. Papers ISSCC*, Feb. 1964, pp. 32–33.
- [11] R. J. Widlar, "New developments in IC voltage regulators," *IEEE Journal of Solid-State Circuits*, vol. SC-6, no. 1, pp. 2–7, Feb. 1971.
- [12] T. Verster, "P-n junction as an ultralinear calculable thermometer," *Electronics Letters*, vol. 4, no. 9, pp. 175–176, May 1968.
- [13] R. C. Dobkin, "Monolithic temperature transducer," in *Dig. Techn. Papers ISSCC*, Feb. 1974, pp. 126–127.
- [14] G. C. M. Meijer, "Integrated circuits and components for bandgap references and temperature transducers," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, Mar. 1982.
- [15] A. J. M. Boomkamp and G. C. M. Meijer, "An accurate biomedical temperature transducer with on-chip microcomputer interfacing," in *Proc. ESSCIRC*, Sept. 1985, pp. 420–423.
- [16] M. J. S. Smith, L. Bowman, and J. D. Meindl, "Analysis, design, and performance of micropower circuits for a capacitive pressure sensor IC," *IEEE Journal of Solid-State Circuits*, vol. SC-21, no. 6, pp. 1045–1056, Dec. 1986.
- [17] G. C. M. Meijer *et al.*, "A three-terminal integrated temperature transducer with microcomputer interfacing," *Sensors and Actuators*, vol. 18, pp. 195–206, June 1989.
- [18] E. A. Vittoz, "MOS transistors operated in the lateral bipolar mode and their application in CMOS technology," *IEEE Journal of Solid-State Circuits*, vol. SC-18, no. 3, pp. 273–279, June 1983.

- [19] P. Krummenacher and H. Oguey, "Smart temperature sensor in CMOS technology," *Sensors and Actuators*, vol. A21-A23, pp. 636–638, Mar. 1990.
- [20] G. Wang and G. C. M. Meijer, "Temperature characteristics of bipolar transistors fabricated in CMOS technology," *Sensors and Actuators*, vol. 87, pp. 81–89, Dec. 2000.
- [21] F. Fruett and G. C. M. Meijer, *The Piezjunction Effect in Silicon Integrated Circuits and Sensors*. Boston: Kluwer Academic Publishers, May 2002.
- [22] J. F. Creemer, "The effect of mechanical stress on bipolar transistor characteristics," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, Jan. 2002.
- [23] A. Bakker and J. H. Huijsing, "Micropower CMOS temperature sensor with digital output," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 7, pp. 933–937, July 1996.
- [24] K. B. Klaassen, "Digitally controlled absolute voltage division," *IEEE Transactions on Instrumentation and Measurement*, vol. 24, no. 2, pp. 106–112, June 1975.
- [25] R. J. van der Plassche, "Dynamic element matching for high-accuracy monolithic D/A converters," *IEEE Journal of Solid-State Circuits*, vol. SC-11, no. 6, pp. 795–800, Dec. 1976.
- [26] G. C. M. Meijer, G. Wang, and F. Fruett, "Temperature sensors and voltage references implemented in CMOS technology," *IEEE Sensors Journal*, vol. 1, no. 3, pp. 225–234, Oct. 2001.
- [27] "ADT7301 data sheet," Analog Devices Inc., Aug. 2004, www.analog.com.
- [28] "LM92 data sheet," National Semiconductor Corp., Mar. 2005, www.national.com.
- [29] "DS1626 data sheet," Maxim Int. Prod., May 2005, www.maxim-ic.com.
- [30] M. A. P. Pertijs, G. C. M. Meijer, and J. H. Huijsing, "Precision temperature measurement using CMOS substrate PNP transistors," *IEEE Sensors Journal*, vol. 4, no. 3, pp. 294–300, June 2004.
- [31] F. Fruett, G. C. M. Meijer, and A. Bakker, "Minimization of the mechanical-stress-induced inaccuracy in bandgap voltage references," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 7, pp. 1288–1291, July 2003.

Chapter 2

CHARACTERISTICS OF BIPOLAR TRANSISTORS

Bipolar transistors form the core of most smart temperature sensors. This chapter reviews the physics of bipolar transistors and the various effects that determine the temperature dependency of their base-emitter voltage. The bipolar transistors available in standard CMOS processes are described and compared. Their most important non-idealities are discussed, including their sensitivity to processing spread and mechanical stress. The models introduced in this chapter will be used extensively in the rest of the book.

2.1 Introduction

As mentioned in Chapter 1, most smart temperature sensors essentially digitize the thermal voltage kT/q , using the bandgap voltage V_g as reference. To help understand where these voltages come from, the basics of bipolar transistors physics will be reviewed. First, the characteristics of junction diodes will be discussed, along with the non-idealities that prevent the use of such diodes in temperature sensors. Then, building on the same principles, the operation of bipolar transistors will be explained, and their relevant non-idealities will be reviewed.

2.1.1 The Ideal Diode Characteristic

Since the p-n junction diode forms the basic building block of a bipolar transistor, and shows an exponential current-voltage characteristic that is similar to that of a bipolar transistor, it is useful to briefly review its physical backgrounds [1, 2].

When a piece of p-type material is joined together with a piece of n-type material, electrons will diffuse from the n-side to the p-side, and holes will diffuse in the opposite direction. Positive donor ions and negative acceptor ions

are left uncovered near the junction in the n-side and p-side respectively, forming a depletion region. The resulting electric field causes electrons and holes to drift in a direction opposite to the diffusion currents. In thermal equilibrium, there is no net current across the junction, and the drift and diffusion currents balance each other out.

If an external voltage V is applied to a p-n junction, the balance between the drift and diffusion currents is disturbed. For a positive voltage (forward bias), the drift current is reduced in favour of the diffusion current, resulting in net current flowing from the p-side to the n-side and a decrease in the depletion region width. The concentration of minority electrons on the p-side of the depletion region is then given by

$$n_{p\text{-side}} = n_{p0} \exp\left(\frac{qV}{kT}\right), \quad (2.1)$$

where n_{p0} is the equilibrium concentration of electrons. Similarly, the concentration of minority holes on the n-side of the depletion region is given by

$$p_{n\text{-side}} = p_{n0} \exp\left(\frac{qV}{kT}\right), \quad (2.2)$$

where p_{n0} is the equilibrium hole concentration. The exponential dependency in these equations results from the Maxwell-Boltzmann approximation to the Fermi-Dirac probability function. According to this approximation, the concentration of electrons in the conduction band and that of holes in the valence band depends exponentially on the position of the Fermi energy level. Since an external voltage V results in a Fermi level that changes throughout the diode (in contrast with the constant Fermi level of a diode in equilibrium), an exponential dependency on V appears [2].

Far away from the depletion region, the minority-carrier concentrations equal their equilibrium values. As a result, the electron concentration on the p-side decreases away from the depletion region, and the hole concentration on the n-side does the same. These concentration gradients give rise to diffusion currents. The total current flowing through the diode is the sum of these diffusion currents, and can be described by

$$I_D = \frac{qAD_n(n_{p\text{-side}} - n_{p0})}{L_n} + \frac{qAD_p(p_{n\text{-side}} - p_{n0})}{L_p}, \quad (2.3)$$

where D_n and L_n are the diffusion constant and the diffusion length of minority electrons on the p-side, D_p and L_p are the diffusion constant and the diffusion length of minority holes on the n-side, and A is the junction area. Substituting (2.1) and (2.2) in this equation gives the well-known ideal diode characteristic

$$I_D = I_S \left(\exp\left(\frac{qV}{kT}\right) - 1 \right), \quad (2.4)$$

where the saturation current I_S is given by

$$I_S = \frac{qAD_n n_{p0}}{L_n} + \frac{qAD_p p_{n0}}{L_p}. \quad (2.5)$$

The equilibrium electron and hole concentrations n_{p0} and p_{n0} can be expressed in terms of the intrinsic carrier concentration n_i and the acceptor and donor concentrations N_a and N_d , respectively:

$$n_{p0} = \frac{n_i^2}{N_a}, \quad p_{n0} = \frac{n_i^2}{N_d}, \quad (2.6)$$

which leads to the following expression for the saturation current I_S :

$$I_S = qAn_i^2 \left(\frac{D_n}{L_n N_a} + \frac{D_p}{L_p N_d} \right). \quad (2.7)$$

For forward-bias voltages $V \gg kT/q$, the -1 term in (2.4) can be neglected. The voltage drop that develops across the diode when a given bias current I is applied is then

$$V = \frac{kT}{q} \ln \left(\frac{I}{I_S} \right). \quad (2.8)$$

If two bias currents I_1 and $I_2 = pI_1$ are successively applied to a diode, the difference in voltage drop is

$$V_2 - V_1 = \frac{kT}{q} \ln \left(\frac{pI_1}{I_S} \right) - \frac{kT}{q} \ln \left(\frac{I_1}{I_S} \right) = \frac{kT}{q} \ln p. \quad (2.9)$$

This difference is proportional to absolute temperature (PTAT) and, ideally, independent of any processing-related parameters. As such, it seems very useful for temperature sensing. However, as will be discussed in the next section, non-ideal effects that are not modelled in equation (2.4) make most diodes unsuitable for accurate temperature sensing [3].

2.1.2 Non-Idealities of Diodes

The ideal diode characteristic (2.4) is based on the assumption that no generation or recombination of electron-hole pairs takes place in the depletion region. In practice, however, some of the carriers injected across the depletion region will recombine. The resulting recombination current can be described by

$$I_{rec} = I_{r0} \exp \left(\frac{qV}{2kT} \right), \quad (2.10)$$

where I_{r0} depends on the width of the depletion region, the recombination lifetime, and the intrinsic carrier concentration [2].

The total diode current is the sum of (2.4) and (2.10). Because of the factor 2 in (2.10), the two exponential terms cannot be easily combined in a single exponential expression. In practice, the following empirical approximation is often used:

$$I = I_S \left(\exp \left(\frac{qV}{nkT} \right) - 1 \right), \quad (2.11)$$

where n is the so-called non-ideality factor [2]. This approximation is valid only for a limited range of voltages. For large forward-bias voltages, when diffusion dominates, $n \approx 1$, and for low forward-bias voltages, when recombination dominates, $n \approx 2$. In the intermediate region, $1 \leq n \leq 2$.

With this empirical approximation, the PTAT voltage given by equation (2.9) becomes proportional to the non-ideality factor n . Since n depends on I_{r0} , which in turn depends on processing parameters, the PTAT voltage becomes sensitive to processing spread. This prevents the use of practical silicon diodes for accurate temperature sensing [3]. Fortunately, as will be shown in the next section, bipolar transistors behave much more ideal, with a non-ideality factor very close to 1.

2.2 Bipolar Transistor Physics

The relation between the collector current I_C and the base-emitter voltage V_{BE} of a bipolar transistor in its forward-active region is similar to the current-voltage characteristic of a diode. The non-idealities that prevent the use of diodes for accurate temperature sensing, however, are not part of the collector current of a bipolar transistor, but give rise to a base current. The $I_C - V_{BE}$ characteristic therefore follows the ideal exponential behaviour much better than the $I - V$ characteristic of a diode.

In this section, first the ideal characteristic of a pnp transistor will be derived. The discussion is based on a pnp transistor rather than an npn transistor because the transistors available in most CMOS processes are pnp transistors. Then, non-idealities of the $I_C - V_{BE}$ characteristic will be discussed. Finally, it will be shown that in some cases the base current is sufficiently well-behaved for transistors to be used in a ‘diode-connected’ configuration, in which their $I_E - V_{BE}$ characteristic is used.

2.2.1 Sign Conventions

Figure 2.1 shows the sign conventions often used in literature for npn and pnp transistors [4]. The polarities of the voltages and currents are chosen such that positive values correspond to the forward-active region of the transistor. For a pnp transistor, the *emitter-base* voltage is positive in this region. Therefore, one should strictly speaking use ‘ V_{EB} ’ and ‘ ΔV_{EB} ’ when discussing temperature

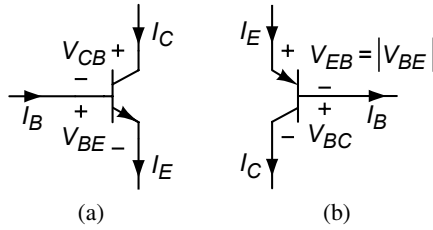


Figure 2.1. Sign conventions used for (a) npn transistors and (b) pnp transistors; positive voltages and currents correspond to the forward-active region.

sensors based on pnp transistors. However, in literature on CMOS temperature sensors and bandgap references (see for instance [5–7]), the symbols ‘ V_{BE} ’ and ‘ ΔV_{BE} ’ are often used instead, so as to be able to use the (more familiar) equations of an npn transistor. In this work, the same convention will be followed. In the equations in the following sections, the symbol ‘ V_{BE} ’ should therefore be strictly speaking be read as $|V_{BE}|$, or V_{EB} .

2.2.2 The Ideal $I_C - V_{BE}$ Characteristic

Consider a pnp transistor in its forward-active region, i.e. with a forward-biased base-emitter junction and a reverse-biased base-collector junction [8]. In this region, the current flow in the base-emitter junction is similar to that in the diode discussed in the previous section: holes are injected into the base region by the emitter, and electrons are injected into the emitter region by the base. The acceptor concentration in the emitter, however, is usually much larger than the donor concentration in base: the base-emitter junction is a so-called one-sided diode [2]. As a result, the diffusion current in the base due to the injection of holes will be much larger than that in the emitter due to the injection of electrons.

The injection of holes results in a minority-carrier concentration $p_{n,em}$ at the emitter-side of the base region that is larger than the equilibrium concentration p_{n0} , but still small compared to the majority-carrier concentration. It depends exponentially on the base-emitter voltage:

$$p_{n,em} = p_{n0} \exp\left(\frac{qV_{BE}}{kT}\right), \quad (2.12)$$

As for the diode discussed in the previous section, this exponential dependency comes from the Boltzmann approximation to the Fermi-Dirac distribution function. As before, the equilibrium concentration p_{n0} can be expressed in terms of the donor concentration N_d in the base and the intrinsic carrier concentration n_i :

$$p_{n0} = \frac{n_i^2}{N_d}. \quad (2.13)$$

Since holes are swept away across the reverse-biased base-collector junction, the minority-carrier concentration at the collector side of the base region is approximately zero. The resulting concentration gradient results in diffusion of minority carriers across the base region. In contrast with the diffusion of holes in the n-side of a diode, the diffusion of holes in the base-region takes place over the relatively short base width W_B . Since this is typically much smaller than the hole diffusion length L_p , the hole concentration decreases approximately linearly from $p_{n,em}$ at the emitter side to zero at the collector side. The corresponding collector current can therefore be described by

$$I_C = \frac{qA\overline{D}_p p_{n,em}}{W_B}, \quad (2.14)$$

where A is the emitter area, W_B is the base width, and \overline{D}_p is the average diffusion constant of holes in the base. By combining (2.12)-(2.14), the collector current can be written as

$$I_C = I_S \exp\left(\frac{qV_{BE}}{kT}\right), \quad (2.15)$$

where the saturation current I_S is given by

$$I_S = \frac{qAn_i^2\overline{D}_p}{W_B N_d}, \quad (2.16)$$

The product $W_B N_d$ is the so-called Gummel number G_B , which expresses the number of impurities per unit area of the base. This Gummel number can also be evaluated as an integral of the doping concentration over the base width if non-uniform doping is used. The effective hole diffusion constant \overline{D}_p is related to the effective hole mobility $\overline{\mu}_p$ via the Einstein relation [2]:

$$\overline{D}_p = \frac{kT}{q}\overline{\mu}_p. \quad (2.17)$$

The saturation current can thus also be expressed as

$$I_S = \frac{kT An_i^2 \overline{\mu}_p}{G_B}. \quad (2.18)$$

2.2.3 Non-Idealities of the $I_C - V_{BE}$ Characteristic

The ideal exponential characteristic described by (2.15) implies that a PTAT voltage can be obtained by biasing a transistor at two collector currents I_{C1} and $I_{C2} = pI_{C1}$, and taking the difference of the resulting base-emitter voltages:

$$\Delta V_{BE} = V_{BE2} - V_{BE1} = \frac{kT}{q} \ln p. \quad (2.19)$$

As for a diode, non-ideal currents need to be accounted for to evaluate how accurate such a PTAT voltage will be in practice.

Two important mechanisms that affect the collector current are the generation of carriers in the base-collector junction, and the diffusion of minority electrons in the collector [2]. These mechanisms result in a leakage current in the base-collector junction that adds to the collector current and would disturb the current ratio p . Fortunately, these currents can be reduced to negligible levels by ensuring that the base-collector voltage is zero.

In that case, however, the assumption that the minority hole concentration at the collector side is zero is not valid anymore. Rather, it will be equal to the equilibrium hole concentration p_{n0} . The hole concentration $p_{n,em}$ at the emitter side due to injection will have to be significantly larger than p_{n0} for (2.15) to be accurate. Replacing $p_{n,em}$ in the diffusion-current equation (2.14) by $(p_{n,em} - p_{n0})$ leads to a more accurate equation for the collector current:

$$I_C = I_S \left(\exp \left(\frac{qV_{BE}}{kT} \right) - 1 \right). \quad (2.20)$$

This equation is identical to the ideal diode equation (2.4). It results in a modification of ΔV_{BE} :

$$\Delta V_{BE} = \frac{kT}{q} \ln \left(\frac{pI_{C1} + I_S}{I_{C1} + I_S} \right) \quad (2.21)$$

$$\simeq \frac{kT}{q} \ln \left(p - \frac{I_S}{I_{C1}}(p - 1) \right), \quad (2.22)$$

which shows that the collector currents I_{C1} and pI_{C1} have to be significantly larger than I_S to obtain an accurate PTAT voltage. Since I_S increases rapidly with temperature (as will be explained later), this requirement is especially relevant at the high end of the operating temperature range.

The collector current cannot be chosen arbitrarily high. A collector current that is too high results in significant self-heating and/or a significant voltage drop across series resistances. Moreover, at a high collector current, the assumption that the minority carrier concentration in the base is low compared to the majority carrier concentration, is not valid anymore. The transistor is then operated in its high-injection region, where $\ln(I_C)$ becomes proportional to $qV_{BE}/(2kT)$ [2]. This effect is illustrated in Figure 2.2a, which shows the logarithm of the collector current as a function of the base-emitter voltage (a so-called Gummel plot).

The reason that a diode can usually not be used to generate an accurate PTAT voltage is the presence of a recombination current with a different temperature dependency than the diffusion current. Recombination also takes place in the base-emitter junction of a bipolar transistor. The crucial difference between a

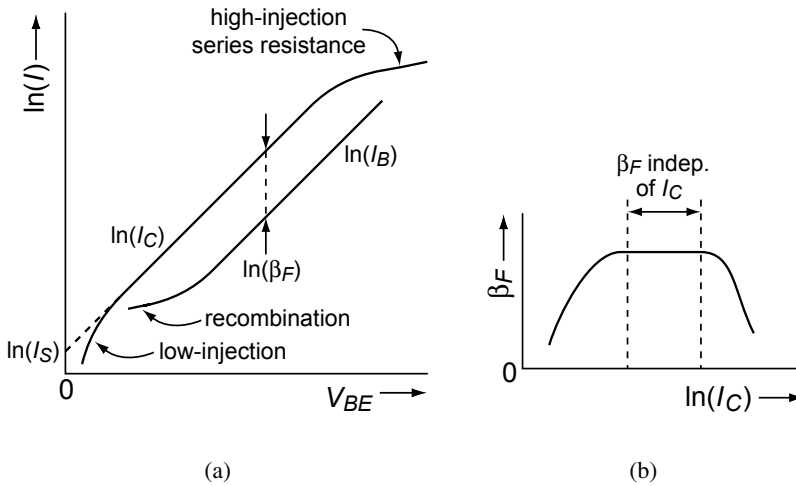


Figure 2.2. (a) The collector current I_C and the base current I_B as a function of the base-emitter voltage V_{BE} , for $V_{BC} = 0$; (b) the associated forward current-gain β_F as a function of the collector current.

diode and a transistor is that the resulting non-ideal components of the emitter current are mainly provided via the base of the transistor. As a result, at low current levels, $\ln(I_B)$ becomes proportional to $qV_{BE}/(2kT)$ (see Figure 2.2a; this effect will be discussed in more detail below). The collector current, in contrast, still follows (2.20) accurately.

In conclusion, in the region where I_C is much larger than I_S and small enough to avoid high injection, the $V_{BE} - I_C$ characteristic of a bipolar transistor is suitable for generating a voltage that is accurately proportional to absolute temperature.

2.2.4 Non-Idealities of the $I_E - V_{BE}$ Characteristic

If the emitter current of a bipolar transistor, rather than its collector current, is set by a current source, the base current has to be taken into account when determining the resulting base-emitter voltage. This is the case in the so-called diode-connected configuration shown in Figure 2.3 [8]. This configuration is especially important for substrate pnp transistors in CMOS technology. These transistors are often connected as a ‘diode’ by grounding their base. This is done because their collector is formed by the substrate, and is therefore not directly accessible. While this connection ensures that the base-collector voltage is zero, it also results in a collector current that is smaller than the applied emitter current:

$$I_C = I_E - I_B = \alpha_F I_E = \frac{\beta_F}{1 + \beta_F} I_E, \quad (2.23)$$

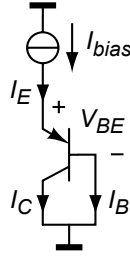


Figure 2.3. A diode-connected pnp transistor.

where α_F is the common-base current-gain, and β_F is the common-emitter current-gain, which equals the ratio of the base current and the collector current. The common-base current-gain α_F expresses how much of the emitter current makes it to the collector, and is ideally one.

To be able to use a diode-connected transistor for generating a difference in base-emitter voltage that is accurately PTAT, as in (2.19), α_F needs to be current-independent, so that a well-defined emitter-current ratio results in the same collector-current ratio. Note that α_F does not necessarily have to be close to one, which would correspond to a high β_F . Rather, α_F (and therefore also β_F) should be constant.

The common-base current-gain α_F can be split up in three factors:

$$\alpha_F = \gamma \alpha_T \delta. \quad (2.24)$$

The factors γ , α_T and δ correspond to the various mechanisms through which base current is formed [2]:

- The emitter injection efficiency γ models the fact that there is a small diffusion current due to electrons that are injected from the base into the emitter, in addition to the hole diffusion current in the base. The associated base current can be described by

$$I_{B1} = \frac{qA n_{iE}^2 \overline{D}_n}{L_n N_a} \exp\left(\frac{qV_{BE}}{kT}\right), \quad (2.25)$$

where n_{iE} is the intrinsic carrier concentration in the emitter, \overline{D}_n and L_n are the average diffusion constant and the diffusion length of minority electrons in the emitter, and N_a is the acceptor concentration in the emitter. The injection efficiency γ can be improved by increasing N_a with respect to the base doping N_a , and by decreasing the base width W_B (which increases I_C while I_{B1} remains the same).

- The base transport factor α_T models the fact that some of the minority holes recombine in the base region as they diffuse from the emitter to the

collector. This effect can be reduced by decreasing the base width W_B . Like the above-mentioned diffusion current, the base current I_{B2} associated with recombination in the base is proportional to $\exp(qV_{BE}/kT)$.

- The recombination factor δ models the combined effects of recombination in the base-emitter junction, surface recombination, and emitter-base channels. These effects are associated with processing defects, and result in a base current component I_{B3} that is proportional to

$$I_{B3} \propto \exp(qV_{BE}/n_E kT), \quad (2.26)$$

where n_E is the so-called low-current forward emission coefficient, which lies between 2 and 4 [9].

The only base-current components that result in a current-dependent α_F , are those that do not share the collector current's relationship to V_{BE} . This is true only for the base-current components associated with δ , which are proportional to $\exp(qV_{BE}/n_E kT)$. At high currents, these components are negligible compared to the other components. At lower currents, however, they become dominant, resulting in a drop in the forward current-gain β_F (see Figure 2.2b).

In conclusion, a diode-connected bipolar transistor can be used to generate an accurate PTAT voltage provided that its recombination factor δ is close to one. The range of emitter currents for which this is the case can be recognized by the fact that the transistor's current gain is independent of the applied emitter current. Not all bipolar transistors have such a range. In some cases, there is insufficient separation between the current level where recombination ceases to dominate the base current, and the current level where high-injection starts [9]. Fortunately, as will be shown in Section 2.4, substrate bipolar transistors, at least in mature CMOS processes, do have such a range, and therefore can be used in a diode-connected configuration to generate an accurate PTAT voltage.

2.3 Temperature Characteristics of Bipolar Transistors

The temperature dependency of the base-emitter voltage V_{BE} of a bipolar transistor in its forward-active region depends on the temperature dependency of the saturation current I_S . Moreover, if the transistor is diode-connected and therefore biased via its emitter, the temperature dependency of the common-base current-gain α_F is also important. This section describes the temperature dependencies of I_S and α_F , and their effect on V_{BE} .

2.3.1 Temperature Dependency of the Saturation Current

To determine the temperature dependency of I_S , the individual terms of equation (2.18) have to be considered :

$$I_S(T) = \frac{kTAn_i^2(T)\overline{\mu}_p(T)}{G_B(T)}. \quad (2.27)$$

The temperature dependency of the intrinsic carrier concentration n_i is given by [10]

$$n_i^2(T) \propto T^3 \exp\left(\frac{-qV_g(T)}{kT}\right), \quad (2.28)$$

where V_g is the bandgap voltage of silicon, which is often assumed to be a linear function of temperature:

$$V_g(T) = V_{g0} - \alpha T, \quad (2.29)$$

In this equation, V_{g0} is the extrapolated bandgap voltage at 0 K [10]. The effective hole mobility $\overline{\mu}_p$ is proportional to T^{-n} , where n is a constant [10]. The Gummel number G_B , finally, is also slightly temperature dependent, because the base width is temperature dependent. Base-width modulation (the Early effect) will be discussed in Section 2.7. For now, G_B will be assumed to be temperature independent.

By combining these temperature dependencies, the saturation current can be written as¹

$$I_S(T) = CT^\eta \exp\left(-\frac{qV_{g0}}{kT}\right), \quad (2.30)$$

where C is a constant, and $\eta = 4 - n$.

Substitution of this expression in (2.15) gives:

$$I_C(T) = CT^\eta \exp\left(\frac{q(V_{BE}(T) - V_{g0})}{kT}\right), \quad (2.31)$$

which can be rewritten as

$$\begin{aligned} V_{BE}(T) &= V_{g0} \left(1 - \frac{T}{T_r}\right) + \frac{T}{T_r} V_{BE}(T_r) \\ &\quad - \eta \frac{kT}{q} \ln\left(\frac{T}{T_r}\right) + \frac{kT}{q} \ln\left(\frac{I_C(T)}{I_C(T_r)}\right), \end{aligned} \quad (2.32)$$

where $V_{BE}(T_r)$ is the base-emitter voltage at a specified reference temperature T_r [4]².

¹This same expression is implemented in SPICE, where the parameters V_{g0} and η are called EG and XTI, respectively [11].

²The reference temperature T_r is called TNOM in SPICE [11].

If values of V_{g0} and η are used that are based on the underlying physics, large differences (several mV) are found between the base-emitter voltage predicted by (2.32) and measurements. Tsividis has shown that this discrepancy is mainly due to the poor modeling of $V_g(T)$ by equation (2.29) [12]. He proposed to use a more accurate model of $V_g(T)$, which reduces the inaccuracy roughly by an order of magnitude.

Meijer, in turn, has shown that even better agreement with measurement data can be obtained if equation (2.32) is used with values of V_{g0} and η that are derived from measured $V_{BE}(T)$ data. Using V_{g0} and η as curve-fitting parameters, the modeling inaccuracy can be reduced to less than 0.1 mV in the temperature range of -20°C to 100°C [4]. The empirical values of V_{g0} and η thus found are often in conflict with the values expected based on physics. For instance, values of η around 4.3 have been found [13], while the physical definition $\eta = 4 - n$ does not allow values larger than 4. This discrepancy results from physical effects, such as the exact temperature dependency of V_g , that are not included in the derivation of (2.32).

As mentioned in Section 1.2, $V_{BE}(T)$ decreases approximately linearly with temperature, with a slope of about $-2\text{ mV}/^\circ\text{C}$. This becomes clear if equation (2.32) is rewritten as the sum of the tangent to the $V_{BE}(T)$ curve at the reference temperature T_r , and a non-linear term $c(T)$ (see Figure 2.4a) [4]:

$$V_{BE}(T) = \underbrace{V_{BE0} - \lambda T}_{\text{tangent at } T=T_r} + c(T), \quad (2.33)$$

where V_{BE0} is the extrapolation of the tangent to $T = 0\text{ K}$,

$$V_{BE0} = V_{g0} - c(0) \quad (2.34)$$

$$= V_{g0} + \frac{kT_r}{q} \left(\eta - \frac{T_r}{I_C(T_r)} \left[\frac{\partial I_C}{\partial T} \right]_{T=T_r} \right), \quad (2.35)$$

λ is the slope of the tangent,

$$\lambda = \frac{V_{BE0} - V_{BE}(T_r)}{T_r}, \quad (2.36)$$

and $c(T)$ is the non-linearity, or curvature,

$$c(T) = \frac{k}{q} \eta \left(T - T_r - T \ln \left(\frac{T}{T_r} \right) \right) + \frac{k}{q} \left(T \ln \left(\frac{I_C(T)}{I_C(T_r)} \right) - (T - T_r) \frac{T_r}{I_C(T_r)} \left[\frac{\partial I_C}{\partial T} \right]_{T=T_r} \right). \quad (2.37)$$

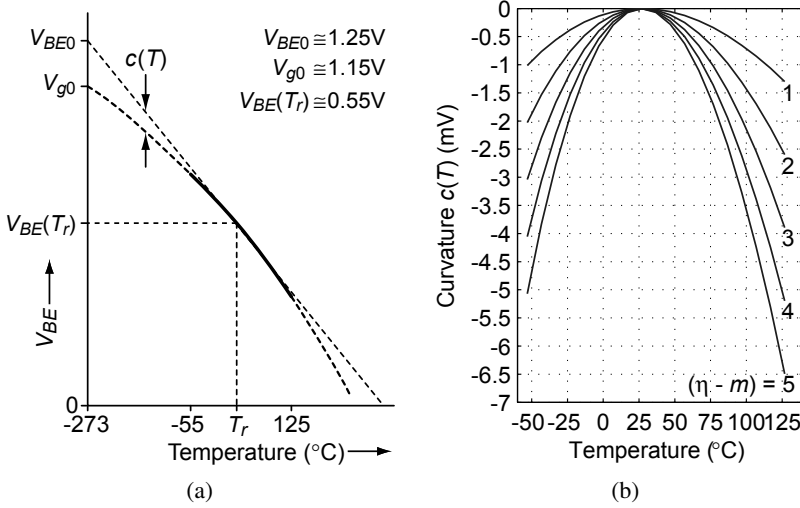


Figure 2.4. (a) Temperature dependency of the base-emitter voltage (the curvature is exaggerated); (b) detail of the curvature for various values of $(\eta - m)$; $T_r = 300\text{K}$.

If the collector current is proportional to a power of T ,

$$I_C(T) = I_C(T_r) \left(\frac{T}{T_r} \right)^m, \quad (2.38)$$

which is the case, for instance, if it is derived from a PTAT voltage ($m = 1$) using a temperature-independent resistor (see Section 2.8.2), the curvature term can be rewritten as

$$c(T) = \frac{k}{q} (\eta - m) \left(T - T_r - T \ln \left(\frac{T}{T_r} \right) \right). \quad (2.39)$$

This expression is plotted in Figure 2.4b for various values of $(\eta - m)$. The curvature is roughly parabolic, and amounts to about 4 mV over the temperature range of -55°C to 125°C for $\eta - m = 3$, which is a typical value for CMOS substrate pnp transistors biased at a PTAT current [13].

2.3.2 Temperature Dependency of the Current Gain

If a bipolar transistor is biased via its emitter, its collector current will be a fraction α_F of the applied emitter current (see equation (2.23)). If α_F varies significantly with temperature, this has to be taken into account when deriving the temperature dependency of V_{BE} . The temperature dependency of α_F can be derived from that of the common-emitter current-gain β_F :

$$\alpha_F(T) = \frac{\beta_F(T)}{1 + \beta_F(T)}, \quad (2.40)$$

where $\beta_F(T)$, in turn, can be derived from the temperature dependency of the various components of the base current. In practice, the following empirical model is often used³

$$\beta_F(T) = \beta_{F0} \left(\frac{T}{T_r} \right)^{X_{TB}}, \quad (2.41)$$

where the nominal current-gain β_{F0} and the temperature exponent X_{TB} are found by fitting the equation to measured data.

Substituting $I_C(T) = \alpha_F(T)I_E(T)$ in (2.32) shows that the temperature dependency of α_F results in an additional term in $V_{BE}(T)$:

$$\begin{aligned} V_{BE}(T) &= V_{BE}(T)|_{\alpha_F=\text{constant}} + \frac{kT}{q} \ln \left(\frac{\alpha_F(T)}{\alpha_F(T_r)} \right) \\ &= V_{BE}(T)|_{\alpha_F=\text{constant}} + \frac{kT}{q} \ln \left(\frac{(1 + \beta_{F0}) \left(\frac{T}{T_r} \right)^{X_{TB}}}{1 + \beta_{F0} \left(\frac{T}{T_r} \right)^{X_{TB}}} \right). \end{aligned} \quad (2.42)$$

This additional term is shown in Figure 2.5a for various values of β_{F0} and X_{TB} . The additional curvature, shown in Figure 2.5b, is only significant for small values of β_{F0} and large values of X_{TB} . If necessary, compensation techniques for finite current-gain can be used to reduce the additional term in V_{BE} to negligible levels (see Section 3.6).

2.4 Bipolar Transistors in Standard CMOS Technology

As mentioned in Section 1.3, low-cost digital CMOS processes only support bipolar transistors as parasitic devices that are a by-product of the MOS transistors. In an n-well CMOS process, they come in two flavours: substrate (vertical) pnp transistors and lateral pnp transistors. Bandgap references and temperature sensors based on both substrate transistors [14, 15] and lateral transistors [16, 17] have been reported. In the following, the characteristics of these transistors will be described, and it will be argued that the substrate pnp transistor should be the device of choice in CMOS temperature sensors [6, 7, 18].

2.4.1 Lateral pnp Transistors

The cross-section of a lateral pnp is shown in Figure 2.6 [16]. This device is essentially a PMOS transistor that is operated as a bipolar transistor. The emitter and the collector are formed by the p^+ source and drain regions, while the base is formed by the n-well. A positive voltage is applied to the gate of the PMOS transistor to prevent the formation of an inversion layer at the surface, and to

³In SPICE, β_{F0} and X_{TB} are called BF and XTB [11].

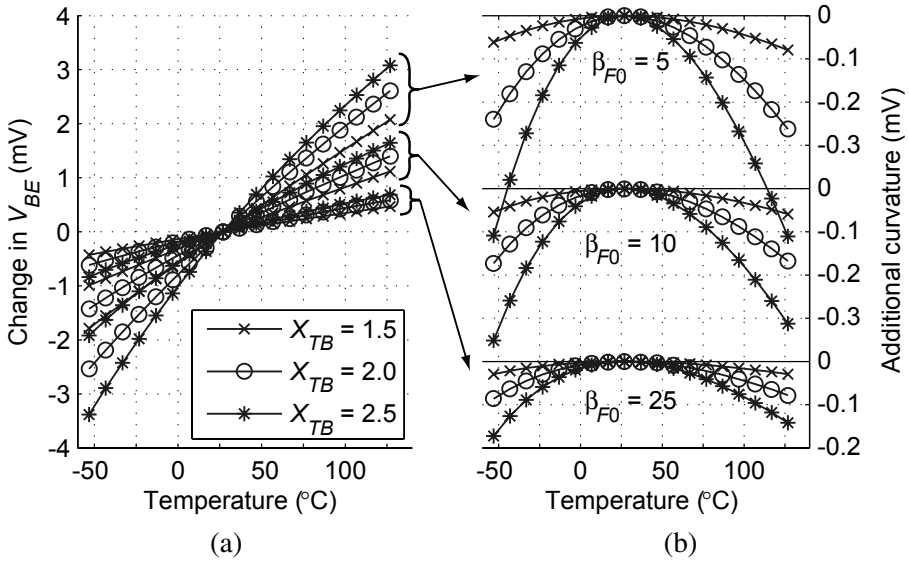


Figure 2.5. (a) Change in the base-emitter voltage due to the temperature dependency of the current gain, for different values of X_{TB} and β_{F0} ; (b) additional curvature as a result of this temperature dependency.

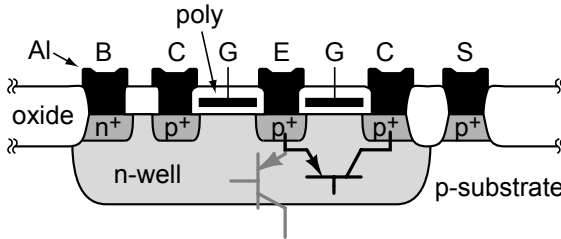


Figure 2.6. Vertical cross-section of a lateral pnp transistor in n-well CMOS technology.

create a depletion layer which pushes the diffusion current of the lateral bipolar transistor beneath the surface. The associated low $1/f$ noise makes lateral pnp transistors attractive for use in low-noise amplifiers and current mirrors [16].

A disadvantage of lateral pnp transistors is that they inevitably have an associated parasitic substrate pnp, formed by the p⁺ source (emitter), the n-well (base) and the substrate (collector). This transistor is drawn in gray in Figure 2.6. It causes a substantial part of the current injected by the emitter to flow vertically into the substrate rather than laterally into the p⁺ collector [16]. The resulting $I_E - V_{BE}$ characteristic is very non-ideal. Lateral pnp transistors therefore have to be biased via their collector, which precludes their use in a

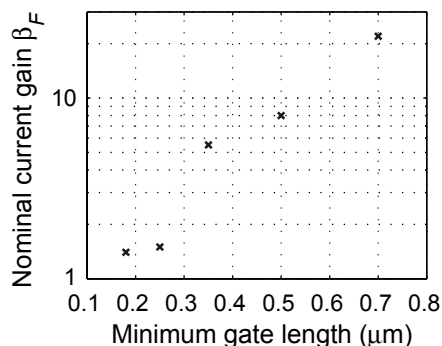


Figure 2.8. Forward current-gain β_F of substrate pnp transistors as a function of minimum gate length, for 0.18 μm , 0.25 μm , 0.35 μm , 0.5 μm and 0.7 μm CMOS processes (data from several design manuals).

substrate transistors are used. Fortunately, modified circuits can be used to employ these transistors in spite of their grounded collector [6, 14, 21]. Such circuits will be discussed in detail in the next chapter.

The base width of substrate pnp transistors is relatively large, as it is mainly determined by the depth of the n-well (typically a few microns). As a result, the common-emitter current-gain β_F of these transistors is very low compared to transistors fabricated in bipolar processes, and also lower than that of lateral pnp transistors in the same process. Moreover, β_F appears to decrease with every new process generation, as illustrated in Figure 2.8. This is due to the higher n-well doping (or retrograde n-well doping) used in denser processes to prevent punch-through [22].

The current flow in substrate pnp transistors is much more one-dimensional than in lateral pnp transistors. As a result, their $I_C - V_{BE}$ characteristic closely follows the ideal exponential behavior over several decades of current [13]. Given the grounded collector of substrate pnp transistors, a crucial question is whether the current gain β_F is independent of the emitter current for a usable range of currents, so that a well-defined emitter-current ratio results in a well-defined collector-current ratio, and hence an accurate ΔV_{BE} . Fortunately, such a region indeed exists: current gains that are almost independent of the emitter current over at least two decades of current have been reported for substrate pnp transistors in both 0.5 μm and 0.7 μm CMOS technology [13]. The relatively current-independent β_F can be explained by the fact that a wide base makes the base-current components proportional to $\exp(qV_{BE}/kT)$ not only large (which makes β_F low), but also dominant compared to the non-ideal components that result in a current-dependent β_F and α_F (see Section 2.2.4).

Another advantage of the wide base of substrate pnp transistors is that variations in the depth of the n-well diffusion or the p^+ implant will have relatively little effect on the transistor's saturation current (see Section 2.5). If the transistor's emitter area is large enough, variations due to lithographic errors will also be small. A lateral pnp transistor, in contrast, will have an effective emitter area that varies with the depth of the p^+ implant. Moreover, lithographic errors will have a much larger effect, because they affect the relatively small base width. Therefore, the relative spread of the saturation current of lateral pnp transistors is expected to be larger than that of substrate pnp transistors in the same process. For lack of statistical data, however, this remains a hypothesis.

A final positive consequence of a wide base is that the base-width modulation due to V_{BE} (the reverse Early effect) will be small. This effect will be discussed in more detail in Section 2.7.

A disadvantage of substrate pnp transistors is their relatively high base resistance and collector resistance (due to the absence of a buried layer) [21]. The voltage drop across these resistances has to be taken into account (see Section 2.7).

Wang has shown that the temperature dependency of the base-emitter voltage of substrate pnp transistors in $0.5\ \mu\text{m}$ and $0.7\ \mu\text{m}$ CMOS technology can be modelled using equation (2.32) with errors less than $\pm 0.1\ \text{mV}$ [13]. This results in maximum errors in the order of $\pm 0.03^\circ\text{C}$ in a smart temperature sensor (as will be shown in the next chapter), which is sufficient even for precision applications.

Given these considerations, substrate pnp transistors are preferred over lateral pnp transistors for implementation of bandgap references and temperature sensors in CMOS technology. The restrictions imposed by their grounded collector and their relatively large resistances can be overcome using the circuit techniques discussed in the next chapter.

2.5 Processing Spread

The reproducibility of the base-emitter voltage V_{BE} is important for the accuracy of smart temperature sensors. Assuming the error contribution of the readout electronics is negligible, this reproducibility determines the initial accuracy of these sensors. Moreover, if trimming is used to adjust temperature errors resulting from spread of V_{BE} , it depends on the nature of this spread how effective this trimming will be. This section describes the effect of spread of the saturation current and spread of the current gain on V_{BE} .

Table 2.1. Spread of the n-well sheet resistance in various CMOS processes [23], along with the corresponding variation in the base-emitter voltage V_{BE} of pnp transistors in these processes, assuming that their saturation current I_S spreads in the same way.

process	# of runs	n-well sheet resistance spread		equivalent V_{BE} spread (at 300 K)	
		min	max	min	max
TSMC 0.18 μm	19	-3.5%	+4.0%	-0.91 mV	+1.02 mV
TSMC 0.25 μm	28	-6.5%	+7.8%	-1.72 mV	+1.91 mV
TSMC 0.35 μm	35	-1.9%	+1.6%	-0.45 mV	+0.40 mV
AMI 0.5 μm	63	-1.8%	+3.4%	-0.47 mV	+0.87 mV
AMI 1.5 μm	75	-13.1%	+11.6%	-3.60 mV	+2.80 mV

2.5.1 Spread of the Saturation Current

If a given transistor has a saturation current that deviates by ΔI_S from the nominal value I_S , its base-emitter voltage can be written as

$$\begin{aligned} V_{BE} &= \frac{kT}{q} \ln \left(\frac{I_C}{I_S + \Delta I_S} \right) \\ &= \frac{kT}{q} \ln \left(\frac{I_C}{I_S} \right) - \frac{kT}{q} \ln \left(1 + \frac{\Delta I_S}{I_S} \right) \end{aligned} \quad (2.43)$$

$$\simeq V_{BE}|_{\Delta I_S=0} - \frac{kT}{q} \frac{\Delta I_S}{I_S} \quad (\Delta I_S \ll I_S). \quad (2.44)$$

To estimate the initial accuracy of V_{BE} (i.e. before trimming), the spread of the elements of equation (2.18),

$$I_S = \frac{kT A n_i^2 \overline{\mu_p}}{W_B N_d}, \quad (2.45)$$

needs to be examined.

- The base doping N_d is subject to tolerances in the production process. For bipolar transistors in an n-well CMOS process (see Section 2.4), the base region is formed by an n-well. An estimate of the spread of the base doping can therefore be obtained from the spread of the n-well sheet resistance [24, 25]. Foundries often specify tolerances up to $\pm 50\%$ for this sheet resistance [22], which would translate to similar variations in I_S , resulting in variations in the order of ± 13 mV in V_{BE} . Practical variations, fortunately, are much smaller. Table 2.1 shows the spread of the n-well sheet resistance obtained from test data of the MOSIS prototyping service for several CMOS processes [23], along with the equivalent variation in V_{BE} .

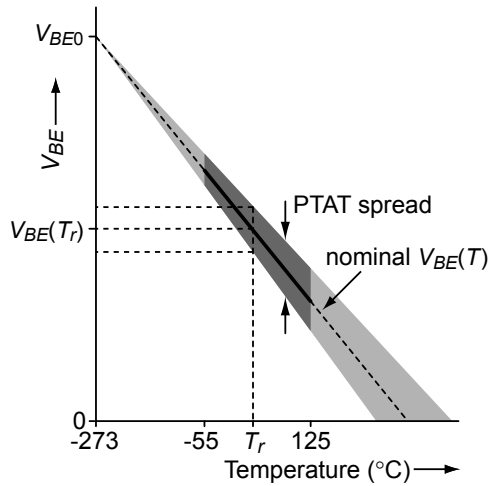


Figure 2.9. Spread of the base-emitter voltage as a result of temperature-independent saturation current spread $\Delta I_S/I_S$ (curvature omitted for clarity).

Especially for modern processes, the expected variations are much smaller than the specified $\pm 50\%$.

- The base width W_B and the emitter area A spread as a result of lithographic errors and variations in the depth of diffusions. For a vertical transistor, the spread of the emitter area is determined by lithography, and can be made negligible by making the emitter area large enough. Lithographic tolerances can be estimated as $\pm 20\%$ of the minimum feature size of the process [22]. The area of a $20\ \mu\text{m} \times 20\ \mu\text{m}$ emitter in a $0.5\ \mu\text{m}$ CMOS process, for example, will then spread by $\pm 1\%$, causing a V_{BE} spread of $\pm 0.25\ \text{mV}$.

The base width is determined by the difference in the depth between the base and the emitter diffusions, and is typically small to optimize the transistor's current gain. Significant spread can therefore be expected. A substrate pnp in CMOS technology, with its wide base, is an exception to this rule (see Section 2.4.2). Exact numbers, however, are not known to the author.

- The intrinsic carrier concentration n_i and the average diffusion constant $\overline{\mu_p}$ change if the transistor is exposed to mechanical stress [26]. This will be discussed in Section 2.6.

It is hard to draw precise conclusions about the initial accuracy of V_{BE} , and therefore about the initial accuracy of smart temperature sensors. The spread of I_S depends on several factors that cannot be controlled by the circuit

designer, and for which statistics are often not available. If a substrate pnp transistor is monitored by a foundry as part of a process control module, the temperature during these measurements is typically not measured accurately enough to obtain useful information about V_{BE} spread (temperature variations of $\pm 1^\circ\text{C}$ already correspond to variations in V_{BE} of $\pm 2\text{ mV}$). The values of Table 2.1 suggest that at least a few mV of V_{BE} spread should be expected. This translates to a best-case initial accuracy of about $\pm 1^\circ\text{C}$ at room temperature.

If a higher accuracy is desired, smart temperature sensors need to be trimmed to correct for V_{BE} spread. Equation (2.43) shows that this spread is proportional to absolute temperature (PTAT), under the assumption that $\Delta I_S/I_S$ is temperature independent. As illustrated in Figure 2.9, such a PTAT spread causes the V_{BE} versus T curve to ‘rotate’ around a fixed point at absolute zero. In the expression for the temperature dependency of V_{BE} , equation (2.32), only $V_{BE}(T_r)$ spreads. The curvature is not affected, and trimming can therefore be performed based on a calibration at only one temperature.

The assumption that $\Delta I_S/I_S$ is independent of temperature is probably quite accurate as far as ΔI_S is the result of spread of the base doping N_d and spread of the dimensions W_B and A . Spread in n_i and $\overline{D_p}$ as a result of mechanical stress, in contrast, results in a temperature-dependent $\Delta I_S/I_S$. Therefore, V_{BE} will still deviate from its nominal value after a PTAT correction. The resulting inaccuracy will be examined in the Section 2.6.

2.5.2 Spread of the Current Gain

If a diode-connected transistor is used for generating V_{BE} , not only spread of the saturation current, but also spread of the common-base current-gain α_F will result in spread of V_{BE} . If this current gain deviates by an amount $\Delta\alpha_F$ from its nominal value α_F , the resulting base-emitter voltage can be written as

$$\begin{aligned} V_{BE} &= \frac{kT}{q} \ln \left(\frac{(\alpha_F + \Delta\alpha_F) I_E}{I_S} \right) \\ &= \frac{kT}{q} \ln \left(\frac{\alpha_F I_E}{I_S} \right) + \frac{kT}{q} \ln \left(1 + \frac{\Delta\alpha_F}{\alpha_F} \right) \end{aligned} \quad (2.46)$$

$$\simeq V_{BE}|_{\Delta\alpha_F=0} + \frac{kT}{q} \frac{\Delta\alpha_F}{\alpha_F} \quad (\Delta\alpha_F \ll \alpha_F). \quad (2.47)$$

This shows that $\Delta\alpha_F$ affects V_{BE} in a similar way as the spread of the saturation current ΔI_S discussed above. Provided that $\Delta\alpha_F/\alpha_F$ is independent of temperature, current-gain spread results in a PTAT spread of V_{BE} that can be trimmed out.

It is, however, not likely that $\Delta\alpha_F/\alpha_F$ is independent of temperature. This can be explained as follows. The physical origin of current-gain spread lies in the spread of the various base-current components discussed in Section 2.2.4.

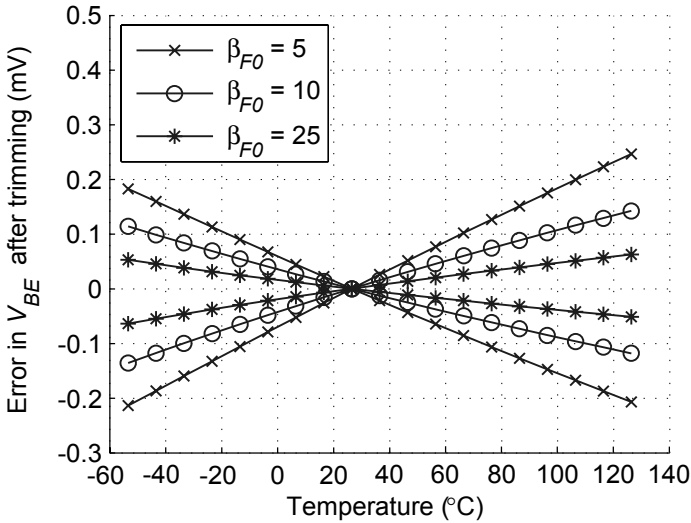


Figure 2.10. Residual error in V_{BE} if errors resulting from $\pm 10\%$ spread of the nominal current-gain β_{F0} are trimmed at 27°C using a PTAT correction; the temperature exponent X_{TB} of the current gain is assumed to be 2.

This causes spread $\Delta\beta_F$ of the nominal common-emitter current-gain β_{F0} . Using $\alpha_F = \beta_F / (1 + \beta_F)$, $\Delta\alpha_F/\alpha_F$ can be written as:

$$\frac{\Delta\alpha_F}{\alpha_F} \simeq \frac{1}{1 + \beta_{F0}} \left(\frac{T}{T_r}\right)^{X_{TB}} \frac{\Delta\beta_F}{\beta_{F0}}, \quad (2.48)$$

where the temperature dependency of β_F , as given by (2.41), has been taken into account. This shows that even if $\Delta\beta_F/\beta_{F0}$ is independent of temperature, $\Delta\alpha_F/\alpha_F$ is not. This means that the resulting spread of V_{BE} will not be PTAT, and can therefore not be trimmed out completely.

Figure 2.10 shows the error in V_{BE} after trimming, for $\Delta\beta_F/\beta_F = \pm 10\%$ and $X_{TB} = 2$. As expected from (2.48), the error in V_{BE} is smaller for larger nominal values of β_{F0} . Unfortunately, β_{F0} of substrate pnp transistors decreases with every new process generation (see Figure 2.8). Therefore, it is important to somehow correct for the errors due to finite current-gain. Techniques for doing this will be discussed in Section 3.6.

2.6 Sensitivity to Mechanical Stress

The characteristics of bipolar transistors are sensitive to mechanical stress. Since die assembly and packaging introduce such stress, it is important to

quantify its effects on the base-emitter voltage, and to identify means of minimizing these effects.

2.6.1 Causes of Mechanical Stress

Mechanical stress is the result of differences in thermal expansion coefficients of the materials of which an integrated circuit is composed. During manufacturing of the die, different materials, such as silicon, silicon oxide and aluminum, are combined at high temperatures. When the die cools down, the difference in expansion between these materials causes local stress on the die, which can result in shifts in component parameters, and mismatches between neighboring components. Similarly, packaging exposes the whole die to stress, as a result of the difference in expansion coefficient between silicon and the packaging materials. The resulting shift in component parameters is referred to as ‘packaging shift’ [22].

Two steps in the packaging process determine the magnitude and nature of packaging stress. First, the die is attached to a lead frame. This is performed at a temperature above the operating temperature range of the chip (a typical temperature for epoxy adhesives is 175 °C). The thermal expansion coefficient of typical die-attach adhesives and lead frames is larger than that of silicon. Therefore, when cooling down to room temperature, the adhesive and lead frame shrink more than the die. As a result, the devices at the surface of the die are exposed to tensile stress [18,27]. For epoxy, typical values for this stress are 35-55 MPa [27]. Larger values should be expected for gold eutectic bonding and solder bonding [18].

After the die has been attached to the lead frame, the chip is encapsulated. If a metal can or ceramic package is used, the die sits inside a cavity and is not exposed to additional stress. Molded plastic packages, in contrast, expose the die to substantial additional stress. Plastic molding takes place at 175 °C. After cooling, the total stress can be either tensile or compressive, with a bias towards compressive [27, 28]. It can have values up to 200 MPa [18]. Moreover, it is not stable under conditions of thermal cycling [29].

The magnitude of the stress in a plastic package can be reduced by coating the die with a mechanically compliant sandwich layer before molding [22]. Abesingha *et al.* have shown a reduction of the packaging shift in a BiCMOS bandgap voltage reference of about 50% [28]. Their reference, which was based on npn transistors, showed a systematic shift of about -5 mV without such a layer, which was reduced to about -2.3 mV using a $15\ \mu\text{m}$ proprietary coating. In addition to this systematic shift, a random shift with a standard deviation of about half the systematic shift was also observed.

Devices that need to match (such as a pair of transistor used for generating ΔV_{BE}) are sensitive to stress gradients, and should therefore be laid-out in a

common-centroid configuration. They should be placed at the center of the die, where the stress gradients are smallest. Care should also be taken to ensure that devices that are supposed to be equal have equal surroundings. Any local stress due to neighboring devices will then affect these devices in the same way. Local stress is often caused by metal interconnect. Metal lines should therefore never be routed across sensitive devices [22].

2.6.2 Stress-Induced Changes in the Saturation Current

The sensitivity of bipolar transistors to mechanical stress is caused by the so-called piezjunction effect. A detailed theoretical and experimental study of this effect can be found in [26] and [18]. Mechanical stress deforms a transistor, changing both its geometry and, through deformation of the crystal structure, the conductivity of the minority carriers in the base. The effect of the resulting changes in the transistor's emitter area A and base width W_B are usually negligible. The change in conductivity, in contrast, is significant, and modifies the saturation current I_S , equation (2.18), through n_i^2 and $\bar{\mu}_p$. The change in I_S depends on the type of transistor and the direction of the current flow, and on the magnitude, type and orientation of the applied stress [18, 26]. Moreover, it also depends on temperature. Fortunately, it hardly depends on the current density, except in the high-injection region. As a result, ΔV_{BE} is insensitive to stress, provided that the current levels are low enough to avoid high-injection. The base-emitter voltage V_{BE} , in contrast, is directly affected by the stress-induced changes in I_S [18].

Packaging causes mainly in-plane stress at the surface of the die, in the range of ± 200 MPa, which, as mentioned above, is tensile (positive) for metal can or ceramic packages, and mostly compressive (negative) for plastic packages. The resulting change in the base-emitter voltage of a substrate pnp transistor is shown in Figure 2.11. The graphs in this figure are calculations based on the empirical model of Fruett [18].

The high non-linearity of the piezjunction effect causes vertical pnp transistors, such as the substrate pnp transistors found in CMOS processes, to be much less sensitive to tensile stress than to compressive stress. The former causes the base-emitter voltage to change by at most $+0.2\text{mV}/-0.1\text{mV}$, while the latter results in changes up to -1.8mV . [18]. This means that much better reproducibility of the base-emitter voltage can be expected in ceramic or metal can packages than in plastic packages.

It is interesting that vertical npn transistors, which are often used for bandgap references and temperature sensors in BiCMOS or bipolar processes, have a much larger sensitivity to mechanical stress than vertical pnp transistors. For compressive stress, the sensitivity is roughly twice as large, while for small tensile stress it can be even up to five times as large. So even if vertical npn transistors are available, vertical pnp transistors are to be preferred [18].

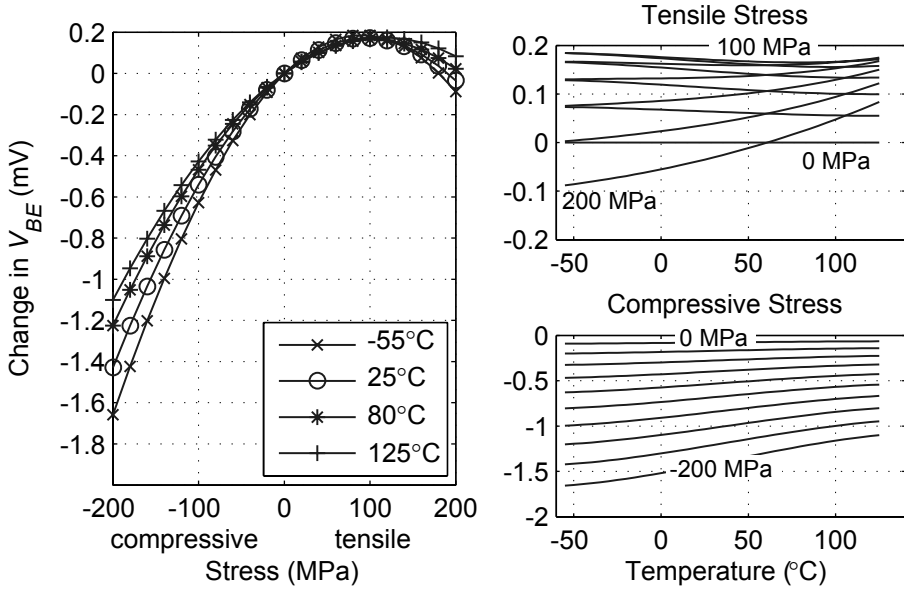


Figure 2.11. Calculated change in the base-emitter voltage of a vertical pnp transistor as a function of mechanical stress in the [110] orientation for different temperatures (left), and as a function of temperature for both tensile and compressive stress in steps of 20 MPa (right); based on extrapolation of data from [18], which covers the temperature range from -10°C to 110°C .

2.6.3 Stress-Induced Changes in the Current Gain

The base-emitter voltage of a diode-connected transistor is not only affected by the stress sensitivity of the saturation current, but also by that of the current gain. A stress-induced change $\Delta\beta_F$ in the current gain modifies V_{BE} as follows:

$$V_{BE} \simeq V_{BE}(T)|_{\Delta\beta_F=0} + \frac{kT}{q} \frac{1}{(1 + \beta_F)} \frac{\Delta\beta_F}{\beta_F} \quad (\Delta\beta_F \ll \beta_F). \quad (2.49)$$

The stress sensitivity of the current gain of vertical pnp transistors has been investigated by Fruett. For in-plane normal stress in the range of -200 MPa to 200 MPa, $\Delta\beta_F/\beta_F$ is less than $\pm 4\%$ [18]. The resulting change in V_{BE} depends on the nominal value of β_F . If, for example, $\beta_F = 5$, the maximum change in V_{BE} is ± 0.17 mV. For compressive stress, this is negligible compared to changes in V_{BE} resulting from stress-induced changes in I_S (see Figure 2.11). For tensile stress, however, the changes are of the same order of magnitude. Compensation techniques for finite current-gain can be used to

make changes in V_{BE} due to the stress sensitivity of β_F negligible (see Section 3.6).

2.7 Effect of Series Resistances and Base-Width Modulation

So far, voltage drop across series resistances associated with the three terminals of a bipolar transistor has been ignored. Also, the base width W_B has been assumed constant, while in practice, it depends on the base-emitter and base-collector voltages (the so-called Early effect). These effects will be discussed in this section.

2.7.1 Series Resistances

Figure 2.12a shows how the effect of ohmic resistances can be included in a circuit model of a diode-connected pnp transistor [9]. The base-emitter and base-collector voltages across the transistor's junctions have been renamed to $V_{B'E'}$ and $V_{B'C'}$, respectively. The externally observed base-emitter voltage V_{BE} is now the sum of the intrinsic base-emitter voltage $V_{B'E'}$ and the voltage drop across the base resistance R_B and the emitter resistance R_E :

$$\begin{aligned} V_{BE} &= V_{B'E'} + I_E R_E + I_B R_B \\ &= V_{B'E'} + I_E \left(R_E + \frac{R_B}{\beta_F + 1} \right). \end{aligned} \quad (2.50)$$

This shows that the base and emitter resistances can be modelled as a single resistance R_S in series with the emitter (Figure 2.12b):

$$R_S = R_E + \frac{R_B}{\beta_F + 1}, \quad (2.51)$$

so that the base-emitter voltage can be expressed as

$$V_{BE} = \frac{kT}{q} \ln \left(\frac{I_{bias}}{I_S} \right) + I_{bias} R_S, \quad (2.52)$$

where the effect of finite current-gain is ignored for simplicity. The series resistance also affects ΔV_{BE} (Figure 2.12c):

$$\Delta V_{BE} = \frac{kT}{q} \ln(pr) + I_{bias} \left(p - \frac{1}{r} \right) R_S, \quad (2.53)$$

where the transistor with the larger emitter area is assumed to be made of a parallel combination of r smaller transistors.

For substrate pnp transistors in CMOS technology, which typically have a low current-gain and a high base resistance, R_S will be dominated by the base

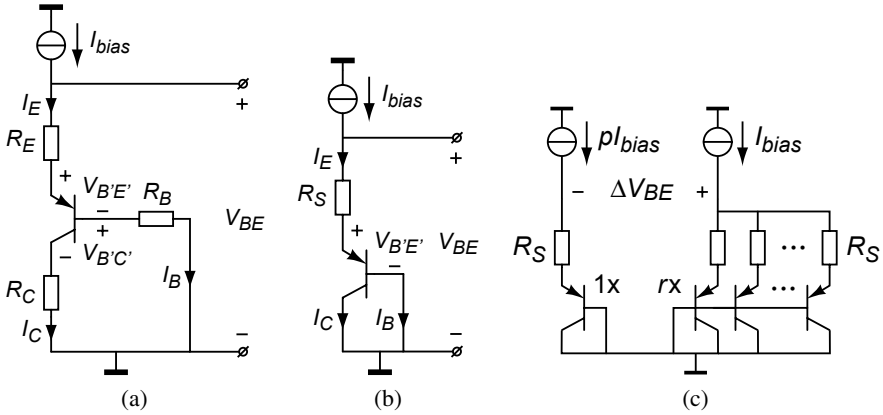


Figure 2.12. (a) Diode-connected pnp with series resistances; (b) equivalent circuit with effective series resistance R_S ; (c) effect of series resistance on ΔV_{BE} .

resistance. For example, for a substrate pnp in $0.7 \mu\text{m}$ CMOS technology with a $10 \mu\text{m} \times 20 \mu\text{m}$ emitter, Wang has reported a base resistance of about 500Ω and a current gain of about 25 at room temperature [13]. The effective series resistance is then about 20Ω . A similar transistor in $0.5 \mu\text{m}$ CMOS technology has a base resistance of about 125Ω and a current gain of 7, giving an effective series resistance of 18Ω [13]. For currents in the μA range, these values lead to a voltage drop of a few tens of μV . While this is usually negligible compared to V_{BE} , it may be significant compared to ΔV_{BE} , and should therefore be taken into account.

Base resistance can be reduced by optimizing the geometry of the transistor, e.g. by using a fingered emitter with base contacts in between the fingers [22]. Such a fingered emitter, however, results in a less one-dimensional device (the fraction of the emitter-base junction that is located at the surface is increased with respect to a circular or square emitter). The associated non-idealities (such as a surface recombination) may degrade the performance of a diode-connected device. For a given transistor geometry, the error due to series resistance can be reduced by reducing the bias current, or, alternatively, by putting multiple devices in parallel. Alternatively, the compensation techniques that will be discussed in Section 3.7 can be used.

2.7.2 Forward Early Effect

Voltage drop across the base resistance R_B and the collector resistance R_C causes the voltage $V_{B'C'}$ across the base-collector junction to be non-zero. A change in the base-collector voltage modifies the width of the depletion layer of the base-collector junction, which in turn changes the effective base width W_B of the transistor. This so-called forward Early Effect causes the collector

current of a bipolar transistor in its forward-active region to depend on the base-collector voltage [2]. This can be included in the model of the $I_C - V_{BE}$ characteristic as follows [9]:

$$I_C = \frac{I_S}{1 + V_{B'C'}/V_{AF}} \exp\left(\frac{qV_{B'E'}}{kT}\right), \quad (2.54)$$

where V_{AF} is the forward Early voltage. This can be rewritten as

$$V_{B'E'} = \frac{kT}{q} \left\{ \ln\left(\frac{I_C}{I_S}\right) + \ln\left(1 + \frac{V_{B'C'}}{V_{AF}}\right) \right\} \quad (2.55)$$

$$\simeq \frac{kT}{q} \left\{ \ln\left(\frac{I_C}{I_S}\right) + \frac{V_{B'C'}}{V_{AF}} \right\} \quad (V_{B'C'} \ll V_{AF}). \quad (2.56)$$

The base-collector voltage of a diode-connected transistor is

$$\begin{aligned} V_{B'C'} &= I_B R_B - I_C R_C \\ &= I_C \left(\frac{R_B}{\beta_F} - R_C \right). \end{aligned} \quad (2.57)$$

Substituting in (2.56) gives

$$V_{B'E'} \simeq \frac{kT}{q} \ln\left(\frac{I_C}{I_S}\right) + I_C \frac{kT}{qV_{AF}} \left(\frac{R_B}{\beta_F} - R_C \right) \quad (2.58)$$

$$= \frac{kT}{q} \ln\left(\frac{I_C}{I_S}\right) + I_E \frac{kT}{qV_{AF}} \left(\frac{\frac{R_B}{\beta_F+1} - \beta_F R_C}{\beta_F + 1} \right). \quad (2.59)$$

This shows that the modification of $V_{B'E'}$ by the forward Early effect can be modelled as a resistor in series with the emitter, which can be added to R_S in equation (2.51) [30]. Since typically $V_{AF} \gg kT/q$ (a typical value for V_{AF} is 100 V, while $kT/q \simeq 25$ mV at $T = 300$ K), this addition to the effective series resistance is usually negligible.

2.7.3 Reverse Early Effect

The base width W_B is not only modulated by the base-collector voltage, but also by the base-emitter voltage. This is the reverse Early effect, which can be incorporated in the $I_C - V_{BE}$ model in a similar way as the forward Early effect [9]:

$$I_C = \frac{I_S}{1 + V_{B'C'}/V_{AF} + V_{B'E'}/V_{AR}} \exp\left(\frac{qV_{B'E'}}{kT}\right), \quad (2.60)$$

where V_{AR} is the reverse Early voltage. Neglecting the forward Early effect, this equation can be rewritten as

$$V_{B'E'} = \frac{kT}{q} \left\{ \ln \left(\frac{I_C}{I_S} \right) + \ln \left(1 + \frac{V_{B'E'}}{V_{AR}} \right) \right\} \quad (2.61)$$

$$\simeq \frac{1}{1 - \frac{kT}{qV_{AR}}} \frac{kT}{q} \ln \left(\frac{I_C}{I_S} \right). \quad (2.62)$$

This shows that the reverse Early effect introduces a multiplicative error in the base emitter voltage. Note that V_{AR} in these equations is *not* the Early voltage of the transistor operated in its reverse active region (base-collector junction forward biased, base-emitter junction reverse biased) [31]. Rather, it should be determined from measurements of the transistor in its forward active region, for instance by measuring the slope of the $\ln(I_C)$ versus V_{BE} characteristic [9].

The multiplicative error in (2.62) is very similar to that modelled by the so-called forward emission coefficient (or non-ideality factor) n_F in many circuit simulators [11], which modifies $V_{B'E'}$ as follows:

$$V_{B'E'} = n_F \frac{kT}{q} \ln \left(\frac{I_C}{I_S} \right). \quad (2.63)$$

Comparing this with (2.62) shows that the reverse Early effect can be modelled with a value of n_F of

$$n_F \simeq \frac{1}{1 - \frac{kT}{qV_{AR}}} \simeq 1 + \frac{kT}{qV_{AR}}, \quad (2.64)$$

which is slightly larger than 1. Deviations from 1 in the order of 0.1% have been reported for CMOS substrate pnp transistors [13]. Surprisingly, also values slightly smaller than 1 have been found. A satisfying explanation for this is not known to the author.

Circuit simulators typically use a constant value for n_F , which means that the temperature dependency in (2.64) is not taken into account. Therefore, the reverse Early effect can be best modelled using a finite value to V_{AR} combined with $n_F = 1$, rather than with an infinite V_{AR} combined with a non-unity n_F [30].

Modeling of the reverse Early effect is especially important in bandgap references [32] and temperature sensors with a voltage output [30]. In a smart temperature sensor, in contrast, the multiplicative error cancels due to the *ratiometric* nature of the sensor: a temperature-dependent voltage is divided by a reference voltage, both of which are based on base-emitter voltages. If all base-emitter voltages contain the same multiplicative error, which can be achieved by operating them at roughly the same current levels, this error cancels in the division, and can therefore be ignored.

Table 2.2. Typical values for the sheet resistance and the first- and second-order temperature coefficients α_{TCR1} and α_{TCR2} of resistors in CMOS technology (compiled from design manuals of several CMOS processes with feature sizes between $0.35\mu\text{m}$ and $0.7\mu\text{m}$).

resistor type	R_{sheet} (Ω/\square)	α_{TCR1} ($\times 10^{-3} \text{ K}^{-1}$)	α_{TCR2} ($\times 10^{-6} \text{ K}^{-2}$)
poly	25 ... 150	-1.1 ... +0.9	0 ... +3
high-ohmic poly	1000 ... 2000	-2.1 ... -0.9	0 ... +5
p ⁺ active	70 ... 100	+1.1 ... +1.9	0 ... +2
n ⁺ active	50 ... 70	+1.4 ... +1.5	0 ... +2
n-well	800 ... 1400	+3.5 ... +5.0	+8 ... +15

2.8 Effect of Variations in the Bias Current

The accuracy of a CMOS smart temperature sensor is not only determined by the characteristics of the bipolar transistors used to generate V_{BE} and ΔV_{BE} , but also by the currents used to bias these transistors. While ΔV_{BE} is unaffected by the magnitude of the bias current (only the ratio p is important, as expressed by equation (2.19)), V_{BE} directly depends on the magnitude of the bias current. On-chip bias currents are usually generated from a bias voltage using a resistor R_{bias} . Assuming that the bias voltage is accurately reproducible (which is the case, for instance, if ΔV_{BE} is used as bias voltage), the reproducibility of the bias current is determined by that of R_{bias} . In this section, the resistors available in a standard CMOS process and their characteristics will be reviewed.

2.8.1 Resistors in Standard CMOS Technology

Table 2.2 lists typical values for the sheet resistance and temperature coefficients for the resistor types available in most CMOS processes. Their properties can be summarized as follows [22]:

- *Polysilicon resistors* are made of the same material that is used for the gates of MOS transistors, except that it lies on field oxide rather than gate oxide, and that it is blocked from silicidation. The temperature coefficient of polysilicon resistors depends on what implants they have been exposed to, and can be both positive and negative [33]. Polysilicon resistors are usually the most linear resistors available, although their voltage dependency is not absolutely zero. In some analog CMOS processes, special high-ohmic poly resistors are available that have a much higher sheet resistance. Due to their low doping, they typically have a large negative temperature coefficient.
- *Shallow diffusion resistors* are made of the shallow p⁺ or n⁺ implants used for constructing the sources and drains of PMOS and NMOS transistors.

In an n-well CMOS process, n^+ resistors lie in the p-substrate, while p^+ resistors lie in an n-well. Both types have positive temperature coefficients, which are higher for lower doping levels [34]. In both cases, the associated pn-junction should be reverse-biased to ensure that the resistor is electrically separated from its environment. This reverse-biased junction introduces leakage currents. Moreover, the bias-dependent depletion layer modulates the effective thickness of the resistor and thus introduces some voltage dependency. For p^+ resistors, these problems can be mitigated by biasing the n-well at the same potential as the resistor.

- *Well resistors* have, as a result of relatively low doping levels, a very high sheet resistance. An even higher sheet resistance can be obtained from so-called pinch resistors, which consist of an n-well with a p^+ implant covering the resistor. As long as the p^+ -n-well junction is kept in reverse bias, the associated depletion layer reduces the thickness of the resistor and, as a result, increases the sheet resistance. The junction of the n-well with the p-substrate has to be kept in reverse bias. As a result of the much lower doping in the n-well, the voltage dependency is much worse than that of n^+ resistors, and the temperature coefficient is larger. An interesting property of n-well and pinch resistors is that they are made of the same material that forms the base of substrate pnp resistors. As a result, there is a positive correlation between the n-well sheet resistance and the saturation current of substrate bipolar transistors [25]. This correlation can possibly be exploited in circuits that compensate for process spread (see Section 3.3.3), but the straightforward use of an n-well resistor as bias resistor is not advisable: due to the positive correlation, the resulting spread of V_{BE} is then expected to be *larger* than in the case of uncorrelated spread of I_S and R_{bias} .
- *Metal interconnect* (usually aluminum) can also be used as a resistor; however, its sheet resistance is usually so low (tens of $m\Omega / \square$) that only small resistors can be made within a reasonable chip area.

In conclusion, well resistors and metal resistors are not suitable for use as bias resistor. The choice between polysilicon resistors and shallow diffusion resistors has to be made based on their temperature dependency, and their sensitivity to processing spread and mechanical stress. These subjects are discussed below.

2.8.2 Temperature Dependency of the Bias Resistor

In Section 2.3, it was assumed that the collector current is proportional to a power of T . In practice, the temperature dependency of the collector current depends on that of the bias voltage and that of the bias resistor. Assuming the bias voltage is proportional to a power of T (which is the case, for instance for

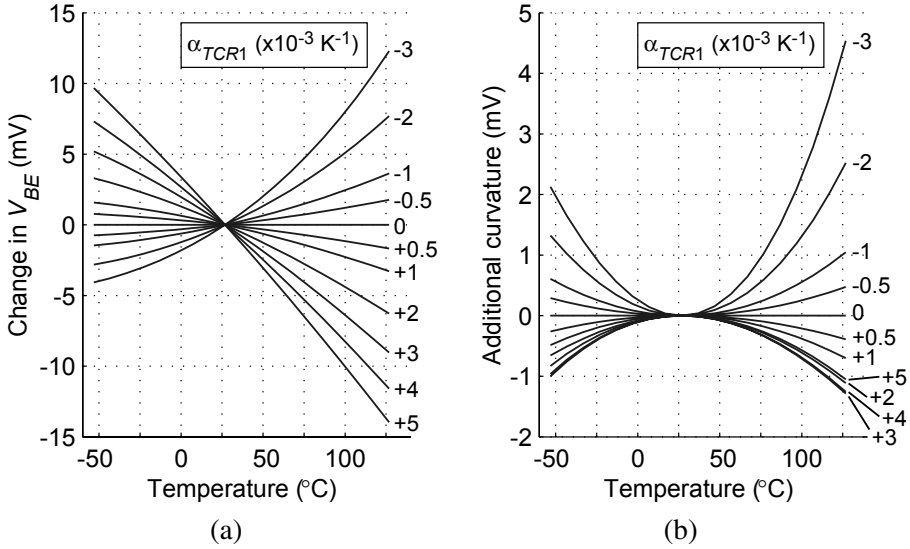


Figure 2.13. (a) Change in the base-emitter voltage due to the temperature dependency of the bias resistor, for different values of its temperature coefficient α_{TCR1} ; (b) additional curvature resulting from this temperature dependency.

a constant or PTAT bias voltage), the collector current can be written as

$$I_C(T) = \frac{V_{bias}(T_r)}{R_{bias}(T)} \left(\frac{T}{T_r} \right)^m. \quad (2.65)$$

The temperature dependency of the bias resistor can be described by a series expansion [11]:

$$R_{bias}(T) = R_0 (1 + \alpha_{TCR1}(T - T_r) + \alpha_{TCR2}(T - T_r)^2), \quad (2.66)$$

where R_0 is the nominal resistance at $T = T_r$, and α_{TCR1} and α_{TCR2} are the first- and second-order temperature coefficients.

Substitution of (2.65) and (2.66) in (2.32) shows that the temperature dependency of R_{bias} results in an additional term in $V_{BE}(T)$:

$$V_{BE}(T) = V_{BE}(T)|_{R_{bias}=\text{constant}} - \frac{kT}{q} \ln (1 + \alpha_{TCR1}(T - T_r) + \alpha_{TCR2}(T - T_r)^2). \quad (2.67)$$

The contribution of the α_{TCR2} term in this equation is usually small. Figure 2.13a shows the change in V_{BE} for several values of α_{TCR1} . The additional term changes both the extrapolated value V_{BE0} , the slope λ , and the curvature.

As shown in Figure 2.13b, the change in the curvature can be substantial for large values of α_{TCR1} . For negative values of α_{TCR1} , the additional curvature

is convex, and thus reduces the overall curvature of V_{BE} . For positive values of α_{TCR1} , the additional curvature is concave, just like that due to $I_S(T)$, and thus increases the overall curvature. For very large positive values of α_{TCR1} , it becomes convex again. Such values, however, do not occur with practical resistors.

2.8.3 Spread of the Bias Resistor

Both the nominal value and the temperature dependency of a bias resistor will vary as a result of processing spread. A change ΔR_{bias} changes the base-emitter voltage as follows:

$$V_{BE} = V_{BE}(T)|_{R_{bias}=\text{constant}} - \frac{kT}{q} \ln \left(\frac{R_{bias} + \Delta R_{bias}}{R_{bias}} \right) \quad (2.68)$$

$$\simeq V_{BE}(T)|_{R_{bias}=\text{constant}} - \frac{kT}{q} \frac{\Delta R_{bias}}{R_{bias}} \quad (\Delta R_{bias} \ll R_{bias}). \quad (2.69)$$

Spread in V_{BE} due to bias resistor spread adds to that due to spread of the saturation current, and thus reduces the initial accuracy of V_{BE} . The value of a resistor can be expressed as a function of the sheet resistance R_{sh} , and the ratio of the resistor's length L and width W [22]:

$$R_{bias} = R_{sh} \frac{L}{W}. \quad (2.70)$$

Both L and W spread due to lithographic tolerances. Since typically $L/W \gg 1$, the relative spread of the W is much larger than that of the L . It can be made negligible by making W several times larger than the minimum width. Spread of R_{bias} will then be determined by that of the sheet resistance, which is determined by the doping and thickness of the resistive layer, and various other parameters, such as the grain structure in the case of polysilicon resistors [22].

Foundries often specify large sheet-resistance tolerances, which would imply a very poor initial accuracy. The variations seen in actual measured process data, such as those of the n-well sheet resistance shown in Table 2.1, are usually smaller. A sheet resistance spread of $\pm 10\%$, for instance, results in a spread of ± 2.5 mV. As discussed in Section 2.5.1, it is hard to draw general conclusions about the resulting initial accuracy of V_{BE} . Typically, trimming will be needed if an accuracy better than a few mV is required (equivalent to better than ± 1 °C in a temperature sensor).

Equation (2.68) shows that the spread in V_{BE} due to bias resistor spread is PTAT provided that $\Delta R_{bias}/R_{bias}$ is independent of temperature. This is the same type of spread resulting from saturation current spread if $\Delta I_S/I_S$ is independent of temperature. As discussed in Section 2.5.1, such PTAT spread can be trimmed out based on a calibration at one temperature. It is therefore important to investigate the validity of the assumption that $\Delta R_{bias}/R_{bias}$ is

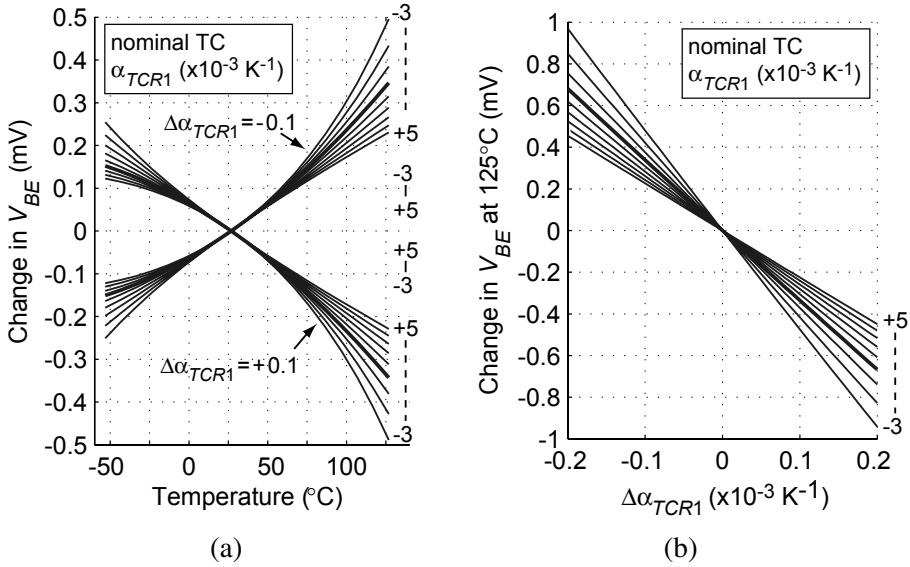


Figure 2.14. Spread in the base-emitter voltage due to spread $\Delta\alpha_{TCR1}$ of the temperature coefficient of the bias resistor: (a) as a function of temperature, for values of α_{TCR1} in the range $-3 \times 10^{-3} \text{ K}^{-1}$ to $+5 \times 10^{-3} \text{ K}^{-1}$ (steps of $1 \times 10^{-3} \text{ K}^{-1}$); (b) at 125°C, as a function of $\Delta\alpha_{TCR1}$, for the same values of α_{TCR1} .

independent of temperature. This is only the case if the temperature dependency of the resistor does not spread.

In practice, the temperature coefficients of resistors depend, among other things, on the doping level [35], and therefore cannot be assumed constant. For mono-crystalline resistors, lower doping levels result in larger positive temperature coefficients [34]. The resistivity of polycrystalline resistors also depends on the doping level, and is negative for low doping levels, and positive for high doping levels [33].

Using (2.67), it can be shown that a change $\Delta\alpha_{TCR1}$ in the first-order temperature coefficient α_{TCR1} modifies V_{BE} as follows

$$V_{BE}(T) \simeq V_{BE}(T)|_{\alpha_{TCR1}=\text{constant}} - \frac{kT}{q} \frac{\Delta\alpha_{TCR1}(T - T_r)}{1 + \alpha_{TCR1}(T - T_r)}, \quad (2.71)$$

where spread of the second-order temperature coefficient α_{TCR2} is ignored. The spread in V_{BE} as a function of temperature is plotted for various values of α_{TCR1} and $\Delta\alpha_{TCR1}$ in Figure 2.14. This spread cannot be corrected for by means of a PTAT trim. It is slightly larger for negative values of α_{TCR1} . Figure 2.14b can be used to find the maximum $\Delta\alpha_{TCR1}$ for a given maximum spread of V_{BE} and a given nominal α_{TCR1} .

If the temperature coefficient spreads too much, calibration at more than one temperature can possibly be avoided if the temperature coefficient is correlated with the nominal value of the resistor. Such a correlation is to be expected, since both the resistivity and the temperature coefficient depend on the doping concentration [34, 36, 37]. Once the correlation has been characterized, the expected value of α_{TCR1} can be calculated from the measured absolute value of R_{bias} at the calibration temperature. An estimate of this value can also be obtained by sheet resistance measurements from a process control module. The sensor can then be trimmed in accordance with the expected value of α_{TCR1} .

2.8.4 Stress-Induced Changes in the Bias Resistor

The sensitivity of resistors to mechanical stress depends on the type of resistor (n- or p-type, mono- or polycrystalline silicon, doping level), its orientation, and on the type and orientation of the applied stress [35]. The value of n-type mono-crystalline resistors decreases as a result of in-plane normal stress. For stress in the range of ± 200 MPa, a decrease up to 5% can be expected [18]. This corresponds to an increase in V_{BE} up to 1.25 mV, which is very significant compared to changes resulting from the stress sensitivity of I_S (see Figure 2.11).

P-type mono-crystalline resistors, such as the p^+ diffusion resistors in CMOS technology, when placed parallel or perpendicular to the wafer flat, can change up to $\pm 10\%$ due to in-plane stress in the range mentioned, resulting in changes up to ± 2.5 mV in V_{BE} . They are relatively insensitive to stress if they are oriented diagonally with respect to the wafer flat (i.e. along the $\langle 100 \rangle$ axis). Design rules, however, do not always allow such an orientation. In that case, a series connection of two orthogonally placed resistors can be used. Fruett and Meijer have shown that this results in a reduction of the sensitivity to stress by a factor 20 [18]. For compressive stress, the resulting changes in V_{BE} are then negligible compared to those due to the stress sensitivity of I_S (see Figure 2.11). For tensile stress, they are of the same order of magnitude as those due to the stress sensitivity of I_S .

The sensitivity of polysilicon resistors to stress is usually a few times smaller than that of mono-crystalline resistors, but it strongly depends on the fabrication process used [35]. For stress in the range mentioned, changes in the order of $\pm 2\%$ should be expected [18, 33]. The associated change in V_{BE} is ± 0.5 mV, which is significant compared to the changes shown in Figure 2.11.

In conclusion, a series combination of two orthogonally placed p-type mono-crystalline resistors is to be preferred as far as sensitivity to mechanical stress is concerned. As discussed in Section 2.8.1, however, such resistors often have high temperature coefficients that may spread, a relatively low sheet resistance that leads to large dimensions, and a high voltage sensitivity due to the reverse-biased junction with the n-well. These properties have to be taken into account when choosing the optimal bias resistor for a given process.

2.9 Conclusions

In this chapter, the physical properties of bipolar transistors and their effect on the voltages V_{BE} and ΔV_{BE} have been discussed. The collector current I_C of a bipolar transistor is accurately proportional to $\exp(qV_{BE}/kT)$, provided that high-current effects (high injection and significant voltage drop across series resistances) and low-current effects (non-negligible leakage currents) are avoided. In that case, a difference in base-emitter voltage ΔV_{BE} can be generated that is accurately PTAT. The $I_E - V_{BE}$ characteristic of some bipolar transistors also follows the ideal exponential characteristic, in spite of the contribution of the base current. This is the case if junction and surface recombination currents are small compared to the collector current. This does not necessarily imply a large current-gain β_F , but rather a range of currents in which β_F is independent of the collector current. If a transistor has such a range, it can be used to generate ΔV_{BE} in a diode-connected configuration.

The temperature dependency of V_{BE} has been described. As a result of the strong temperature dependency of the saturation current I_S , V_{BE} decreases approximately linearly with temperature by about $2 \text{ mV} / ^\circ\text{C}$, from an extrapolated value at absolute zero which is related to the silicon bandgap voltage. The non-linearity, or curvature, of V_{BE} over the military temperature range amounts to a few millivolts. It has been shown that for diode-connected transistors with a small β_F , the temperature dependency of β_F needs to be taken into account, as it results in an increased curvature of V_{BE} . The temperature dependency of the bias current has to be taken into account as well. If this current is generated using a resistor with a positive temperature coefficient (TC), the overall curvature increases, while a negative TC results in a decrease in curvature.

In standard n-well CMOS technology, two flavours of bipolar transistors can be made: lateral pnp transistors and substrate pnp transistors. A disadvantage of substrate pnp transistors is that their collector is formed by the substrate and is hence connected to ground. They follow the ideal exponential characteristic however much better than lateral bipolar transistors. This is because the current flow in a lateral pnp transistor is much less one-dimensional than that in a substrate pnp, causing its parameters to depend on the current level. The temperature dependency of the base-emitter voltage of substrate pnp transistors closely follows the ideal model. Moreover, they have a current region in which β_F is current-independent, so that they can be used in a diode-connected configuration. These advantages make substrate pnp transistors the device of choice for implementation of smart temperature sensors in CMOS technology.

Processing spread causes variations in the saturation current I_S of bipolar transistors. As far as these variations are independent of temperature, they result in PTAT spread of V_{BE} , which can be trimmed out based on a calibration at one temperature. Variations in the bias current due to spread of the nominal value of the bias resistor also result in PTAT spread of V_{BE} . The initial accuracy

of V_{BE} (i.e. before trimming) depends on processing tolerances. Foundries often specify large tolerances, but n-well sheet resistance data suggest that a spread of several mV should be expected. For temperature sensors, this means that trimming will be needed to obtain inaccuracies below roughly $\pm 1^\circ\text{C}$. The difference ΔV_{BE} is insensitive to variations of the saturation current and the bias current.

Like processing spread, mechanical stress also changes the saturation current of bipolar transistors, and thus V_{BE} . Such stress is the result of the difference in thermal expansion coefficients between the die and packaging materials. Substrate pnp transistors are relatively insensitive to tensile stress, which is the dominant type of stress in ceramic and metal can packages. The associated changes in V_{BE} are less than ± 0.2 mV. Their sensitivity to compressive stress, which is dominant in low-cost plastic packages, is much larger, and causes changes up to -2 mV. Bias resistors are also affected by mechanical stress, and the smallest sensitivity can be obtained using orthogonally placed p^+ resistors, or poly resistors.

Due to its temperature dependency, spread of β_F also results in non-PTAT spread of V_{BE} . Therefore, a residual spread remains after PTAT trimming, which becomes larger than ± 0.1 mV for nominal values β_F below 10. The low values of β_F found in modern CMOS processes call for compensation techniques that eliminate this residual spread. Such techniques will be described in Section 3.6.

Spread in the temperature coefficient (TC) of the bias resistor, finally, also results in non-PTAT spread. This spread becomes larger than ± 0.1 mV for spread of the first-order temperature coefficient larger than $\pm 0.025 \cdot 10^{-3} \text{K}^{-1}$. A bias resistor with a low TC-spread should therefore be used (which does not necessarily imply a low TC). Correlations between TC and sheet resistance can possibly be used to predict the TC from resistance measurements. Thus, a calibration at more than one temperature can be avoided even for relatively large TC spread.

The parasitic series resistances associated with the base and the emitter can be modelled as a lumped resistance in series with the emitter. In substrate pnp transistors, the contribution of the base resistance is dominant. Voltage drop across this resistance has to be taken into account when evaluating the accuracy of V_{BE} and ΔV_{BE} , and has to be compensated for if it is too large (see Section 3.7). It has been shown that the effect of the collector resistance of a diode-connected transistor, as a result of the forward Early effect, is a slight increase in the series resistance. This increase is negligible for the high Early voltages found in substrate pnp transistors. The reverse Early effect causes a positive gain error in ΔV_{BE} in the order of 0.1%. Fortunately, an identical gain error affects V_{BE} , so that it cancels in the ratiometric measurement in a smart temperature sensor.

The next chapter discusses how V_{BE} and ΔV_{BE} can be generated accurately using substrate pnp transistors and combined to perform ratiometric temperature measurement. The models introduced in this chapter will be used to evaluate the temperature errors associated with the above-mentioned non-idealities. Various compensation techniques for these non-idealities will be introduced.

References

- [1] S. M. Sze, *Semiconductor devices, physics and technology*. New York: John Wiley & Sons, 1985.
- [2] D. A. Neamen, *Semiconductor Physics and Devices: Basic Principles*, 3rd ed. New York: McGraw-Hill, 2003.
- [3] T. Verster, "P-n junction as an ultralinear calculable thermometer," *Electronics Letters*, vol. 4, no. 9, pp. 175–176, May 1968.
- [4] G. C. M. Meijer, "Integrated circuits and components for bandgap references and temperature transducers," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, Mar. 1982.
- [5] M. Tuthill, "A switched-current, switched-capacitor temperature sensor in 0.6- μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 7, pp. 1117–1122, 1998.
- [6] A. Bakker and J. H. Huijsing, *High-Accuracy CMOS Smart Temperature Sensors*. Boston: Kluwer Academic Publishers, 2000.
- [7] G. C. M. Meijer, G. Wang, and F. Fruett, "Temperature sensors and voltage references implemented in CMOS technology," *IEEE Sensors Journal*, vol. 1, no. 3, pp. 225–234, Oct. 2001.
- [8] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*. Chichester, England: John Wiley & Sons, 2001.
- [9] I. E. Getreu, *Modeling the Bipolar Transistor*. Amsterdam, The Netherlands: Elsevier, 1976.
- [10] J. W. Slotboom and H. C. de Graaff, "Measurements of bandgap narrowing in Si bipolar transistors," *Solid-State Electronics*, vol. 19, pp. 857–862, Oct. 1976.
- [11] T. Quarles, A. R. Newton, D. O. Pederson, and A. Sangiovanni-Vincentelli, *SPICE3 Version 3f3 User's Manual*, University of California, Berkeley, CA, 1993. [Online]. Available: <http://bwrc.eecs.berkeley.edu/Classes/IcBook/SPICE/>
- [12] Y. P. Tsividis, "Accurate analysis of temperature effects in $I_C - V_{BE}$ characteristics with application to bandgap reference sources," *IEEE Journal of Solid-State Circuits*, vol. SC-15, no. 6, pp. 1076–1084, Dec. 1980.
- [13] G. Wang and G. C. M. Meijer, "Temperature characteristics of bipolar transistors fabricated in CMOS technology," *Sensors and Actuators*, vol. 87, pp. 81–89, Dec. 2000.

- [14] Y. P. Tsividis and R. W. Ulmer, "A CMOS voltage reference," *IEEE Journal of Solid-State Circuits*, vol. SC-13, no. 6, pp. 774–778, Dec. 1978.
- [15] A. Bakker and J. H. Huijsing, "Micropower CMOS temperature sensor with digital output," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 7, pp. 933–937, July 1996.
- [16] E. A. Vittoz, "MOS transistors operated in the lateral bipolar mode and their application in CMOS technology," *IEEE Journal of Solid-State Circuits*, vol. SC-18, no. 3, pp. 273–279, June 1983.
- [17] P. Krummenacher and H. Oguey, "Smart temperature sensor in CMOS technology," *Sensors and Actuators*, vol. A21-A23, pp. 636–638, Mar. 1990.
- [18] F. Fruett and G. C. M. Meijer, *The Piezjunction Effect in Silicon Integrated Circuits and Sensors*. Boston: Kluwer Academic Publishers, May 2002.
- [19] D. MacSweeney, K. G. McCarthy, A. Mathewson, and B. Mason, "A SPICE compatible subcircuit model for lateral bipolar transistors in a CMOS process," *IEEE Transactions on Electron Devices*, vol. 45, no. 9, pp. 1978–1984, Sept. 1998.
- [20] F. G. O'Hara, J. J. H. van den Biesen, H. C. de Graaff, W. J. Kloosterman, and J. B. Foley, "MODELLA – a new physics-based compact model for lateral p-n-p transistors," *IEEE Transactions on Electron Devices*, vol. 39, no. 11, pp. 2553–2561, Nov. 1992.
- [21] B.-S. Song and P. R. Gray, "A precision curvature-compensated CMOS bandgap reference," *IEEE Journal of Solid-State Circuits*, vol. SC-18, no. 6, pp. 634–643, Dec. 1983.
- [22] A. Hastings, *The art of analog layout*. New Jersey: Prentice Hall, 2001.
- [23] (2005) MOSIS wafer electrical test data and SPICE model parameters. [Online]. Available: <http://www.mosis.org/test/>
- [24] R. W. Dutton and D. A. Divekar, "Bipolar models for statistical IC design," in *Process and device modeling for integrated circuit design*, F. van de Wiele *et al.*, Eds. Addison-Wesley, 1977, pp. 461–517.
- [25] J. Michejda and S. K. Kim, "A precision CMOS bandgap reference," *IEEE Journal of Solid-State Circuits*, vol. SC-19, no. 6, pp. 1014–1021, Dec. 1984.
- [26] J. F. Creemer, "The effect of mechanical stress on bipolar transistor characteristics," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, Jan. 2002.
- [27] H. Ali, "Stress-induced parametric shift in plastic packaged devices," *IEEE Transactions on Components, Packaging and Manufacturing Technology—Part B: Advanced Packaging*, vol. 20, no. 4, pp. 458–462, Nov. 1997.
- [28] B. Abesingha, G. A. Rincón-Mora, and D. Briggs, "Voltage shift in plastic-packaged bandgap references," *IEEE Transactions on Circuits and Systems—Part II: Analog and Digital Signal Processing*, vol. 49, no. 10, pp. 681–685, Oct. 2002.
- [29] G. C. M. Meijer, "Concepts for bandgap references and voltage measurement systems," in *Analog Circuit Design*, J. H. *et al.*, Ed. Boston: Kluwer Academic Publishers, 1996, pp. 243–268.

- [30] M. A. P. Pertijs, G. C. M. Meijer, and J. H. Huijsing, "Precision temperature measurement using CMOS substrate PNP transistors," *IEEE Sensors Journal*, vol. 4, no. 3, pp. 294–300, June 2004.
- [31] B. L. Hart, "Remarks on the emission coefficient of a bipolar transistor," *Proceedings of the IEEE*, vol. 69, no. 5, May 1981.
- [32] A. van Staveren, C. Verhoeven, and A. van Roermund, "The influence of the reverse Early effect on the performance of bandgap references," *IEEE Transactions on Circuits and Systems—Part I: Fundamental Theory and Applications*, vol. 43, no. 5, pp. 418–421, May 1996.
- [33] E. Luder, "Polycrystalline silicon-based sensors," *Sensors and Actuators*, vol. 10, no. 1-2, pp. 9–23, Sept. 1986.
- [34] P. Norton and J. Brandt, "Temperature coefficient of resistance for p- and n-type silicon," *Solid-State Electronics*, vol. 21, pp. 969–974, 1978.
- [35] S. Middelhoek and S. A. Audet, *Silicon Sensors*. Londen: Academic Press, 1989.
- [36] F. D. King, J. Shewchun, D. A. Thompson, H. D. Barber, and W. A. Pieczonka, "Polycrystalline silicon resistors for integrated circuits," *Solid-State Electronics*, vol. 16, pp. 701–708, 1973.
- [37] W. A. Lane and G. T. Wrixon, "The design of thin-film polysilicon resistors for analog IC applications," *IEEE Transactions on Electron Devices*, vol. 36, no. 4, pp. 738–744, Apr. 1989.

Chapter 3

RATIOMETRIC TEMPERATURE MEASUREMENT USING BIPOLAR TRANSISTORS

The output of a smart temperature sensor is a digital representation of its temperature. This requires a *ratiometric* temperature measurement: the ratio of a temperature-dependent voltage and a reference voltage is determined using an analog-to-digital converter. This chapter describes how substrate bipolar transistors in CMOS technology can be used to accurately generate these voltages.

3.1 Introduction

Section 1.2 described how temperature can be measured by combining the base-emitter voltages of two substrate pnp transistors. The difference in base-emitter voltage ΔV_{BE} between the two transistors was used to generate a voltage proportional to absolute temperature (PTAT). This was combined with the base-emitter voltage V_{BE} of one of the transistors to generate a bandgap reference voltage. The ratio of the PTAT voltage and the reference voltage was then determined by an analog-to-digital converter, in order to obtain a digital representation of temperature.

In this section, such ratiometric temperature measurement will be looked at in more detail. The effects of errors in ΔV_{BE} and V_{BE} will be derived, so that the impact of non-idealities on the overall accuracy can be estimated. The results will be used extensively later on in this chapter and in the following chapters.

3.1.1 Combining V_{BE} and ΔV_{BE}

The principal voltages in the configuration discussed in Section 1.2 are a difference in base-emitter voltages ΔV_{BE} , and an absolute base-emitter voltage V_{BE} . In principle, these voltages can be generated using two transistors, as was

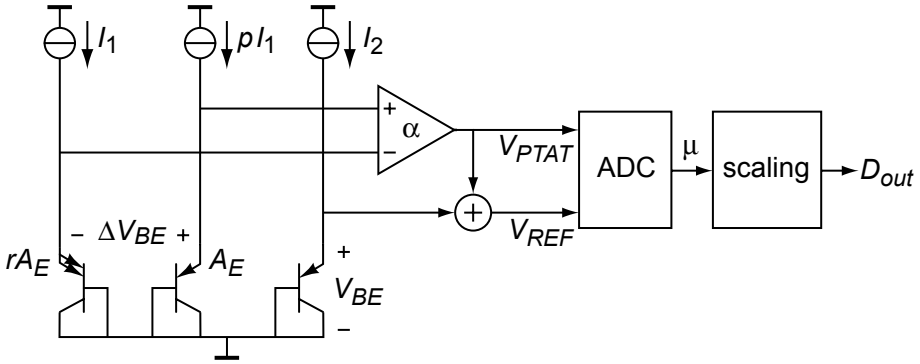


Figure 3.1. Operating principle: three diode-connected pnp transistors are used to generate ΔV_{BE} and V_{BE} . ΔV_{BE} is amplified and combined with V_{BE} to provide an input and reference to an analog-to-digital converter (ADC).

shown in Figure 1.2, or even using a single transistor to which different bias currents are successively applied. To separate the accuracy issues associated with ΔV_{BE} from those associated with V_{BE} , three transistors are used in the conceptual smart temperature sensor shown in Figure 3.1.

Two of these transistors have a well-defined current-density ratio $1 : pr$, resulting from a $1 : p$ bias-current ratio and a $r : 1$ emitter-area ratio. As a result, the difference in their base-emitter voltages is PTAT:

$$\Delta V_{BE} = \frac{kT}{q} \ln \left(\frac{pI_1}{I_S} \right) - \frac{kT}{q} \ln \left(\frac{I_1}{rI_S} \right) = \frac{kT}{q} \ln(pr), \quad (3.1)$$

where I_1 is the unit bias current, and I_S is the saturation current of the smaller transistor. This shows that ΔV_{BE} , in principle, only depends on the ratios p and r and is independent of the absolute bias current and saturation current. In Section 3.2, it will be shown that these ratios can be made accurate by design. For typical values, i.e. $3 < pr < 16$, the sensitivity of ΔV_{BE} is in the order of $0.1 - 0.25 \text{ mV}/^\circ\text{C}$.

The third transistor is used to generate V_{BE} , which depends on the absolute value of the saturation current I_S and the bias current I_2 :

$$V_{BE} = \frac{kT}{q} \ln \left(\frac{I_2}{I_S} \right). \quad (3.2)$$

Its extrapolated value at 0 K is about 1.2 V (related to the silicon bandgap energy), from where it decreases almost linearly by typically $-2 \text{ mV}/^\circ\text{C}$ (Figure 3.2). As discussed in the previous chapter, V_{BE} is affected by various non-idealities. Processing spread causes variations in both I_2 and I_S . In Section

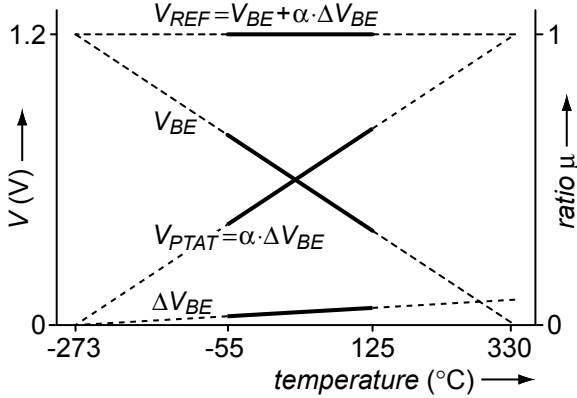


Figure 3.2. Temperature dependency of the various voltages in Figure 3.1.

3.3, it will be shown how I_2 can be generated so as to minimize the effects of this spread. Trimming techniques to compensate for spread will be introduced in Section 3.4. The base-emitter voltage is also slightly non-linear. Correction techniques for this so-called curvature will be described in Section 3.5. For now, V_{BE} will be assumed to be an ideal linear function of temperature, as shown in Figure 3.2.

A nominally temperature-independent reference voltage V_{REF} can be generated by compensating the decrease of V_{BE} with an amplified ΔV_{BE} :

$$V_{REF} = V_{BE} + \alpha \cdot \Delta V_{BE}. \quad (3.3)$$

The gain factor α should be chosen in such a way that the temperature coefficients of V_{BE} and $\alpha \cdot \Delta V_{BE}$ have an equal magnitude and opposite signs:

$$-\frac{\partial V_{BE}}{\partial T} = \alpha \frac{\partial(\Delta V_{BE})}{\partial T}, \quad (3.4)$$

which implies

$$\alpha = -\frac{\partial V_{BE}}{\partial T} \frac{q}{k \ln(pr)}. \quad (3.5)$$

With $\partial V_{BE}/\partial T \simeq -2 \text{ mV}/^\circ\text{C}$ and typical current-density ratios pr between 3 and 16, the value of α ranges between 8 and 20.

A digital temperature reading can be obtained using an analog-to-digital converter (ADC) that takes $V_{PTAT} = \alpha \cdot \Delta V_{BE}$ as input and V_{REF} as reference (see Figure 3.1). The output μ of the ADC will then be

$$\mu = \frac{V_{PTAT}}{V_{REF}} = \frac{\alpha \cdot \Delta V_{BE}}{V_{BE} + \alpha \cdot \Delta V_{BE}} \quad (3.6)$$

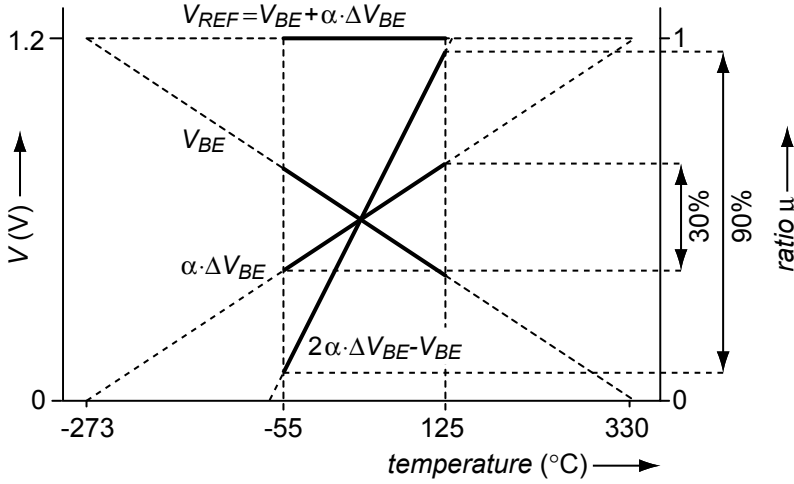


Figure 3.3. More efficient use of the ADC's dynamic range by using a more complicated combination of V_{BE} and ΔV_{BE} .

(see Figure 3.2). Since V_{REF} is temperature-independent, μ will be PTAT. A final digital output D_{out} in degrees Celsius can be obtained by linear scaling:

$$D_{out} = A \cdot \mu + B, \quad (3.7)$$

where $A \simeq 600$ K (since μ goes from 0 to 1 over a temperature range of about 600 K), and $B \simeq -273$ K.

While the transfer obtained with (3.6) is simple, it only uses about 30% of the of the ADC, since the extremes of the operating range correspond to $\mu \simeq 1/3$ and $\mu \simeq 2/3$. Figure 3.3 shows how the dynamic range can be used more efficiently at the expense of a more complicated combination of V_{BE} and ΔV_{BE} [1, 2]. The output of the ADC μ' is then

$$\mu' = \frac{2\alpha \cdot \Delta V_{BE} - V_{BE}}{V_{BE} + \alpha \cdot \Delta V_{BE}}. \quad (3.8)$$

With this more efficient combination, 90% of the dynamic range is used rather than 30%. Thus, the required resolution of the ADC is reduced by a factor of three.

3.1.2 Error Budgeting

As discussed in the previous chapter, V_{BE} and ΔV_{BE} are affected by various non-idealities (process spread, curvature, series resistances, finite current-gain, etc.). As a result, practical values of V_{BE} and ΔV_{BE} will deviate from their ideal values. The readout circuitry will introduce errors as well. The overall

temperature inaccuracy will be the combined result of all these errors. Therefore, a fraction of the maximum overall temperature error has to be allocated to each of the error sources.

When making such an error budget, it is important to distinguish between systematic errors and random errors [3]. Systematic errors are the same for all sensors, and can in principle be corrected by design. Random errors, in contrast, vary from sensor to sensor, and are zero on average. If they are significant, they have to be corrected by means of calibration and trimming. An example of a systematic error is the curvature of V_{BE} (see Section 2.3). Errors due to packaging-induced stress (see Section 2.6) are partially systematic (the so-called packaging shift), and partially random. Examples of errors that are usually purely random are the quantization error of the ADC (which will be discussed in Chapter 4), and offset and mismatch errors in the readout circuitry (which will be discussed in Chapter 5).

When specifying random errors, it is important to express the so-called confidence level. The statement that a random error is $\pm 0.1^\circ\text{C}$ is fairly meaningless, unless the probability that the error lies in that interval is specified. A more exact statement would be that a random error is less than $\pm 0.1^\circ\text{C}$ for 99.7% of the sensors. Alternatively, the standard deviation σ of the distribution can be specified. For a normal distribution, the probability that a random error lies within $\pm\sigma$ is 68.3%, that it lies within $\pm 2\sigma$ is 95.5%, and that it lies within $\pm 3\sigma$ is 99.7% [3].

3.1.3 Errors in V_{BE} , ΔV_{BE} and α

Many errors that will be discussed in this chapter and the following chapters can easily be expressed as an error in V_{BE} , an error in ΔV_{BE} , or an error in the gain α . Therefore, it is useful to calculate the temperature error that corresponds to errors in these values. The sensitivity of the digital output D_{out} to errors in V_{BE} , ΔV_{BE} and α can be calculated from equations (3.6) and (3.7), by differentiating D_{out} to V_{BE} , ΔV_{BE} and α (the use of the alternative transfer function (3.8), incidentally, gives the same results). This results in the following expressions:

$$S_{V_{BE}}^{D_{out}}(T) = \frac{\partial D_{out}}{\partial V_{BE}} = A \frac{\partial \mu}{\partial V_{BE}} \simeq -\frac{T}{V_{REF}}, \quad (3.9)$$

$$S_{\Delta V_{BE}}^{D_{out}}(T) = \frac{\partial D_{out}}{\partial(\Delta V_{BE})} = A \frac{\partial \mu}{\partial(\Delta V_{BE})} \simeq \frac{A - T}{V_{REF}} \alpha, \quad (3.10)$$

$$S_{\alpha}^{D_{out}}(T) = \frac{\partial D_{out}}{\partial \alpha} = A \frac{\partial \mu}{\partial \alpha} \simeq \frac{T}{\alpha} \left(1 - \frac{T}{A}\right), \quad (3.11)$$

where the approximation $\mu \simeq T/A$ has been used.

These sensitivities are shown in Figure 3.4 as a function of temperature. As the sensitivity to errors in ΔV_{BE} also depends on the value of α , the figure

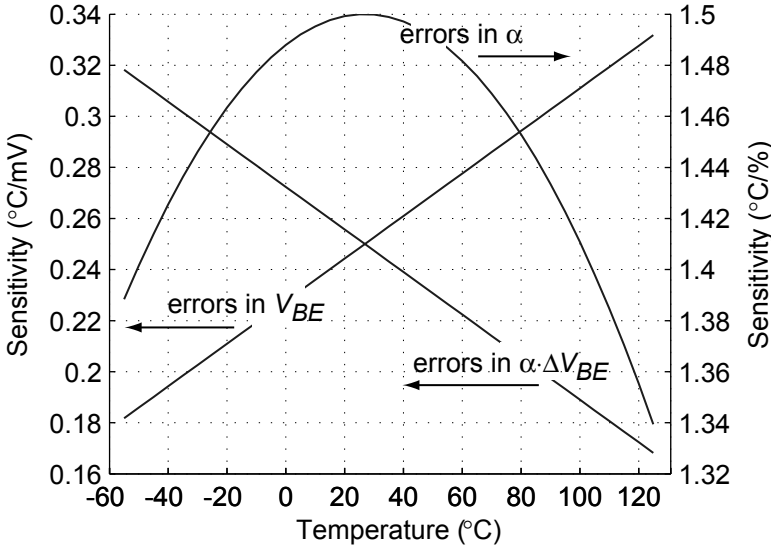


Figure 3.4. Sensitivity of the digital output D_{out} to errors in V_{BE} and $\alpha \cdot \Delta V_{BE}$, and to relative errors in the gain α .

shows $S_{\Delta V_{BE}}^{D_{out}}/\alpha$, i.e. the sensitivity to errors in $\alpha \cdot \Delta V_{BE}$, rather than $S_{\Delta V_{BE}}^{D_{out}}$. Suppose, for example, that α is 10. A 0.1 mV error in ΔV_{BE} then corresponds to a 1 mV error in $\alpha \cdot \Delta V_{BE}$. As can be seen in Figure 3.4, this results in a worst-case error of 0.32 °C at -55 °C. Similarly, a 1 mV error in V_{BE} results in a worst-case error of 0.33 °C at 125 °C. A 1% error in α corresponds to a worst-case error of 1.5 °C at room temperature.

The calculated sensitivities can be used to translate a maximum error contribution ΔT to a maximum error in V_{BE} , ΔV_{BE} or α :

$$|V_{BE} - V_{BE,ideal}| < (3 \text{ mV} / ^\circ\text{C}) \cdot \Delta T, \quad (3.12)$$

$$|\Delta V_{BE} - \Delta V_{BE,ideal}| < \frac{3 \text{ mV} / ^\circ\text{C}}{\alpha} \cdot \Delta T, \quad (3.13)$$

$$\left| \frac{\alpha - \alpha_{ideal}}{\alpha_{ideal}} \right| < \left(\frac{2}{3} \% / ^\circ\text{C} \right) \cdot \Delta T, \quad (3.14)$$

where the worst-case sensitivities have been used. Tighter bounds can be found by taking the temperature dependency in the sensitivities (3.9)–(3.11) into account.

3.1.4 PTAT Errors in V_{BE}

A class of errors that deserves some extra attention are PTAT errors in V_{BE} . As described in the previous chapter, variations in I_S and R_{bias} due to processing

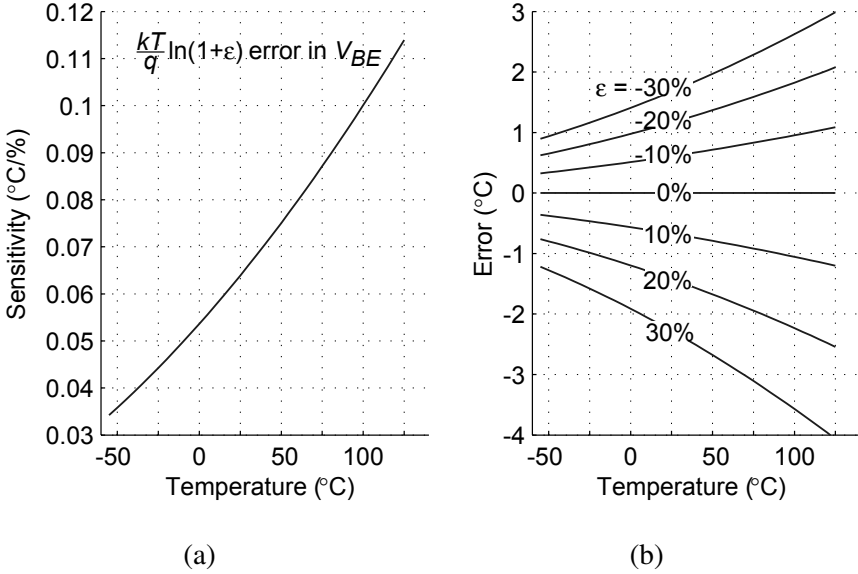


Figure 3.5. (a) Sensitivity of the digital output D_{out} to PTAT errors in V_{BE} ; (b) temperature errors resulting from PTAT errors in V_{BE} .

spread, among other non-idealities, result in such errors (see Figure 2.9). These errors can be written as

$$V_{BE} - V_{BE,ideal} = \frac{kT}{q} \ln(1 + \epsilon) \simeq \frac{kT}{q} \epsilon, \quad (3.15)$$

where ϵ is, for instance, the relative error in I_S or R_{bias} . The sensitivity of D_{out} to these errors is

$$S_{\epsilon}^{D_{out}} = S_{\epsilon}^{V_{BE}} \cdot S_{V_{BE}}^{D_{out}} = \frac{kT}{q} \cdot \frac{T}{V_{REF}}. \quad (3.16)$$

This sensitivity is shown in Figure 3.5a, which shows that PTAT errors have a roughly three times larger affect at the high end of the temperature range than at the low end.

The calculated sensitivity can be used to find the maximum PTAT error in V_{BE} for a given error contribution ΔT . The worst-case value is found at 125 °C:

$$|\epsilon| < (8.7\%/^{\circ}\text{C}) \cdot \Delta T. \quad (3.17)$$

This shows, for instance, that the combined spread of I_S and R_{bias} should be smaller than $\pm 8.7\%$ if a temperature error of $\pm 1^{\circ}\text{C}$ is to be obtained without trimming. The data presented in Table 2.1 suggest that this is feasible for some processes.

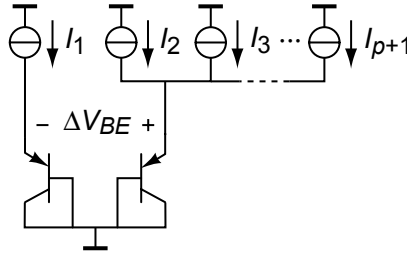


Figure 3.6. Generating a 1 : p bias-current ratio using $p + 1$ identical current sources.

Figure 3.5b shows the temperature error for various values of ε . For large values of ε , the linear approximation of the logarithm in equation (3.15) is not accurate. As a result, negative values of ε result in larger errors than positive values. This figure shows, for instance, that a combined spread of I_S and R_{bias} of $\pm 30\%$ (which is what might be expected from specified process tolerances) leads to temperature errors between $+3^\circ\text{C}$ and -4°C at the high end of the temperature range.

3.2 Generating an Accurate Current-Density Ratio

In this section, errors in ΔV_{BE} resulting from transistor mismatch are discussed. It is shown how dynamic element matching can be applied to reduce these errors.

3.2.1 Errors due to Mismatch

To generate an accurate PTAT voltage, two transistors have to be operated at a well-defined current-density ratio. Such a ratio can be obtained by using two transistors with a different emitter area and/or a different bias current, as shown in Figure 3.1. Ideally, the difference in base-emitter voltages ΔV_{BE} is then (see (3.1))

$$\Delta V_{BE} = \frac{kT}{q} \ln(pr), \quad (3.18)$$

where p is the ratio of the bias currents, and r is the ratio of the emitter areas. Assuming for the moment that $r = 1$, i.e. that two identical transistors are used (or, alternatively, a single transistor which is successively operated at two different bias currents), the accuracy of ΔV_{BE} will depend on that of the current ratio p .

To obtain an accurate current ratio, p must be an integer number [4]. The larger current source can then be constructed from a parallel combination of p identical copies of the smaller current source, as shown in Figure 3.6. Mismatch

Δp between these current sources affects ΔV_{BE} as follows:

$$\Delta V_{BE} = \frac{kT}{q} \ln(p + \Delta p) \quad (3.19)$$

$$\simeq \frac{kT}{q} \ln(p) \cdot \left(1 + \frac{\Delta p}{p \ln p}\right) \quad (\Delta p \ll p). \quad (3.20)$$

The absolute error in ΔV_{BE} can then be written as

$$\Delta V_{BE} - \Delta V_{BE}|_{\Delta p=0} = \frac{kT}{q} \frac{\Delta p}{p}. \quad (3.21)$$

The associated temperature error at the output D_{out} can be found by multiplying this expression by the sensitivity $S_{\Delta V_{BE}}^{D_{out}}$, which is given by (3.10). Since this sensitivity decreases with temperature, and the error (3.21) increases with temperature, the error is largest in the middle of the temperature range, $T = 300$ K, where the sensitivity is $\alpha \cdot 0.25$ K/mV.

If, for instance, $\Delta p/p = 0.1\%$, which is a typical value that can be expected from a careful common-centroid layout of the current sources [4], the error in ΔV_{BE} at $T = 300$ K is $26 \mu\text{V}$. If $p = 10$, then $\alpha \simeq 10$, and the sensitivity is 2.5 K/mV. The temperature error due to mismatch is then 0.065 K.

The upper curves in Figure 3.7 show the temperature error as a function of $\Delta p/p$, for various values of p . If temperature errors well below 0.1 K are required, better matching than 0.1% is needed, even for higher values of p . As will be shown below, dynamic element matching (DEM) can then be used to average out mismatch errors.

Mismatch Δr between the two transistors results in errors in ΔV_{BE} in the same way as mismatch between the current sources.

An easy way to eliminate mismatch between the transistors is to use a single transistor, which is successively biased at a small current and a p times larger current. The switching between these bias currents has to be done fast enough to ensure that the temperature does not change significantly during this procedure. A disadvantage of this approach is that the readout circuitry has to accurately subtract two base-emitter voltages, whereas in the case of two transistors, ΔV_{BE} is instantaneously available as a differential voltage.

3.2.2 Dynamic Element Matching

Errors due to mismatch between the current sources and transistors can be reduced by dynamically interchanging them [5]. This technique is referred to as dynamic element matching (DEM) [6]. If, for instance, a $1 : 3$ bias current ratio is generated using 4 current sources, each of these 4 sources can be used to generate the unit current. This results in 4 possible ΔV_{BE} 's, each of which will have an error due to mismatch. The *average* error, however, will be almost zero.

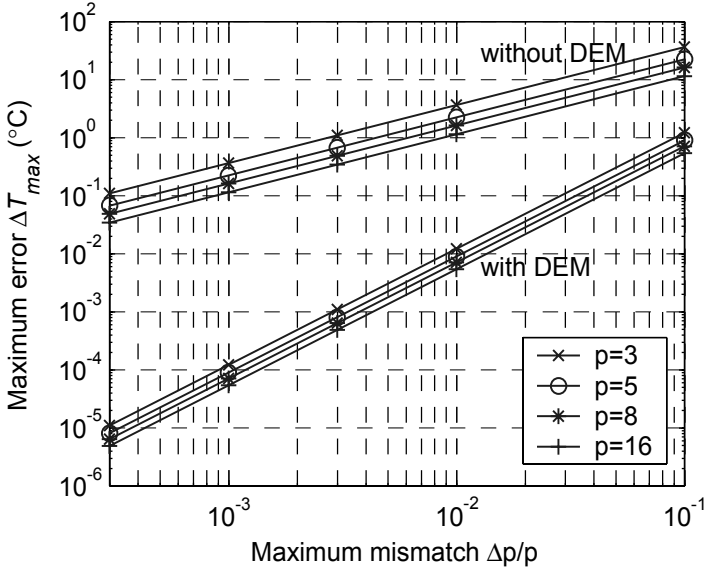


Figure 3.7. Maximum error in ΔV_{BE} as a result of bias-current-source mismatch $\Delta p/p$, expressed as a temperature error at 300K, without dynamic element matching (DEM) and with DEM (see Appendix A).

Dynamic Element Matching of Current Sources

Figure 3.8 shows how dynamic element matching of the bias current sources can be implemented [7]. The current generated by each of the $p + 1$ current sources can be directed either to transistor Q_1 or to transistor Q_2 , which are assumed to be identical for the moment. If current I_j ($1 \leq j \leq p + 1$) is directed to Q_1 , while the other currents are directed to Q_2 , the resulting ΔV_{BE} is:

$$\Delta V_{BE,j} = \frac{kT}{q} \ln \left(\frac{\sum_{i \neq j} I_i}{I_j} \right) = \frac{kT}{q} \ln (p + \Delta p_j), \quad (3.22)$$

where $\Delta p_j/p$ is the mismatch between current I_j and the average of the other currents. The average of the $p + 1$ possible ΔV_{BE} 's is

$$\Delta V_{BE,avg} = \frac{1}{p+1} \sum_{j=1}^{p+1} \Delta V_{BE,j}. \quad (3.23)$$

It can be shown that the first-order errors cancel in this average. A residual second-order error remains, which is bounded as follows (see Appendix A):

$$\left| \Delta V_{BE,avg} - \Delta V_{BE} \Big|_{\Delta p=0} \right| < \frac{1}{2} \frac{kT}{q} \left(\frac{\Delta p}{p} \right)^2, \quad (3.24)$$

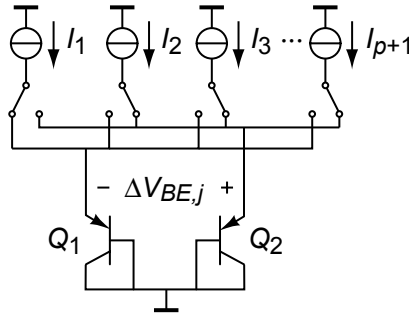


Figure 3.8. Dynamic element matching of bias current sources to generate an accurate ΔV_{BE} .

where $\Delta p/p$ is the worst-case mismatch between the currents, i.e. $|\Delta p_j/p| < \Delta p/p$ for all j .

If, for instance, $p = 10$ and $\Delta p/p = 1\%$, the residual error in ΔV_{BE} at $T = 300$ K is $1.3 \mu\text{V}$. This corresponds to a temperature error of at most 0.003 K (using, as before, $S_{\Delta V_{BE}}^{D_{out}}(300 \text{ K}) \simeq 2.5 \text{ K/mV}$).

The lower curves in Figure 3.7 show the temperature errors that correspond to the bound given by (3.24) as a function of $\Delta p/p$, for various values of p . To obtain temperature errors well below 0.1 K, matching in the order of 1% is sufficient. Such matching can easily be obtained with standard layout techniques.

Dynamic Element Matching of Bipolar Transistors

In principle, mismatch between the transistors Q_1 and Q_2 can be averaged out in the same way (Figure 3.9). A set of $r + 1$ transistors can be used, each of which can be switched as unit transistor, while the others are switched in parallel. Combined with the $p + 1$ DEM steps needed for the current sources, a total of $(r + 1)(p + 1)$ DEM steps is then needed to average out first-order mismatch errors.

A problem with the circuit of Figure 3.9 is, however, that mismatch between the switches in series with the transistors may result in errors in the average ΔV_{BE} . Only if their on-resistances match, the voltage drop across these switches is a pure common-mode voltage.

If two transistors with a nominally equal emitter area are used (that is, $r = 1$), the mismatch between the transistors can be completely eliminated without having any switches directly in series with the emitters. In fact, the circuit of Figure 3.8 can then be used. Suppose that transistors Q_1 and Q_2 have an emitter area ratio $1 : (1 + \Delta r)$. If, during a given step j in the DEM of the current

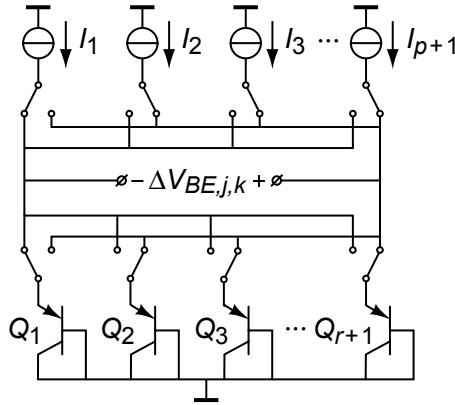


Figure 3.9. Dynamic element matching of both bias current sources and transistors.

sources, current I_j is first directed to Q_1 , the resulting ΔV_{BE} is

$$\Delta V_{BEA,j} = \frac{kT}{q} \ln \left(\frac{p + \Delta p_j}{1 + \Delta r} \right).$$

If current I_j is then directed to Q_2 , ΔV_{BE} changes polarity and becomes

$$\Delta V_{BEB,j} = -\frac{kT}{q} \ln ((p + \Delta p_j) (1 + \Delta r)).$$

By taking the *difference* between these two ΔV_{BE} 's, the error due to Δr cancels completely:

$$\Delta V_{BEA,j} - \Delta V_{BEB,j} = 2 \frac{kT}{q} \ln (p + \Delta p_j).$$

The first-order errors due to current-source mismatch can then be removed, as before, by averaging $p + 1$ of these differences. The complete DEM process now takes $2(p + 1)$ steps, from which an overall average is obtained which is free of first-order mismatch errors.

3.2.3 Errors due to Finite Output Impedance

Practical current sources have a finite output impedance R_o , which causes their output current to depend on the output voltage. This results in an error in the current ratio p , even if dynamic element matching is applied to eliminate mismatch-related errors. In Figure 3.6, the voltage at the output of current source I_1 is ΔV_{BE} lower than that at the output of the other current sources. As a result, the output current of current source I_1 will be slightly higher than

the output current I of the other sources:

$$I_1 = I + \frac{\Delta V_{BE}}{R_o}. \quad (3.25)$$

This results in an error in the ratio p :

$$\frac{\Delta p}{p} \simeq -\frac{\Delta V_{BE}}{IR_o} = -\frac{kT \ln p}{qIR_o} = -\frac{kT \ln p}{qV_A}, \quad (3.26)$$

where V_A is the equivalent Early voltage of the current sources. The absolute error in ΔV_{BE} can then be found using equation (3.21):

$$\Delta V_{BE} - \Delta V_{BE}|_{\Delta p=0} = -\left(\frac{kT}{q}\right)^2 \frac{\ln p}{V_A}. \quad (3.27)$$

This analysis shows that the finite output impedance of the current sources results in a ΔV_{BE} that is systematically too low. As before, the associated temperature error can be calculated by multiplying (3.27) by the sensitivity of the digital output D_{out} to errors in ΔV_{BE} , which is given by (3.10). Due to the T^2 term in (3.27), the temperature error is worst at the high end of the temperature range.

Suppose, for example, that $p = 10$. The gain α is then roughly 10. The sensitivity at the high end of the temperature range, $T = 400$ K, is then 1.7 K / mV. A maximum error ΔT of 0.01 K then corresponds to a maximum error in ΔV_{BE} of 6 μ V. From equation (3.27), using $T = 400$ K, it can be calculated that this maximum corresponds to a minimum Early voltage of 460 V. This shows that cascaded current sources will certainly be required.

3.3 Generating an Accurate Bias Current

In Section 2.8, it has been shown that the base-emitter voltage V_{BE} of a bipolar transistor is sensitive to variations in its collector current, such as those resulting from the temperature dependency, stress sensitivity and processing spread of the resistor used in generating this current. This section discusses how to design bias circuits so as to minimize the associated temperature errors.

3.3.1 Structure of Bias Circuits

As described in Section 2.8, the bias current used for generating V_{BE} is typically derived from a bias voltage V_{bias} using a resistor R_{bias} :

$$I_{bias} = \frac{V_{bias}}{R_{bias}}. \quad (3.28)$$

An important question is then how V_{bias} should be generated. In a sense, this is a chicken-and-egg problem: on-chip bias currents are generated from an

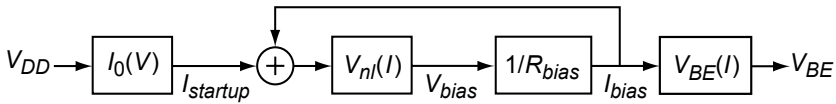


Figure 3.10. Block diagram of a bias circuit.

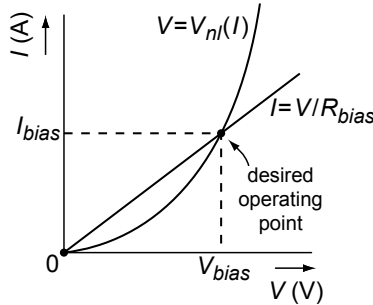


Figure 3.11. Load characteristic of a bias circuit.

on-chip voltage, while a bias current is required to generate a voltage, so which comes first?

The answer lies in startup circuits, which typically generate a startup current from the supply voltage. This is illustrated in Figure 3.10, which shows a block diagram of a typical bias circuit. The bias current I_{bias} is fed back into a non-linear current-to-voltage (I-V) converter $V_{nl}(I)$, which produces the voltage V_{bias} . This, in turn, is converted into a current using a resistor. As illustrated in the load characteristic shown in Figure 3.11, the possible operating points of such a system can be found graphically as the intersections of $V = V_{nl}(I)$ and $I = V/R_{bias}$. The startup current ensures that the undesired operating point where the bias current is zero is avoided.

Since the startup current depends on the supply voltage, care has to be taken to prevent cross-sensitivity of the generated bias current to the supply voltage. This can be done, for instance, by switching off the startup current once the bias circuit has reached its desired operating point [8]. Alternatively, a cascade of bias circuits can be used, so that the total supply rejection becomes the product of that of the individual circuits. A first bias circuit then generates a non-critical bias current (which can be slightly supply dependent), while a second bias circuit, started by the first, generates the supply-independent bias current that is used to generate V_{BE} .

The characteristics of the generated bias current (such as its spread and temperature dependency) depend on the implementation of the non-linear

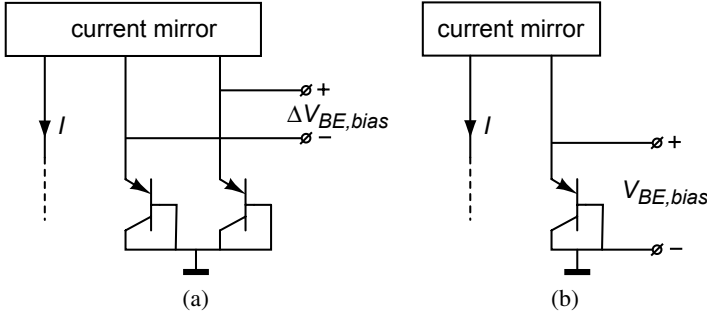


Figure 3.12. Current-to-voltage conversion in a bias circuit: (a) based on a difference in base-emitter voltage; (b) based on a single base-emitter voltage.

I-V converter. In CMOS bias circuits, MOS transistors are often used for this purpose. Since this introduces extra sensitivity to process spread (e.g. through variations in the threshold voltage), such I-V converters will not be considered further here.

A pair of bipolar transistors, as shown in Figure 3.12a, forms a better implementation of the I-V converter. The resulting bias voltage is then

$$V_{bias} = \Delta V_{BE,bias} = \frac{kT}{q} \ln \left(\frac{m I_{bias}}{I_{bias}} \right) = \frac{kT}{q} \ln m, \quad (3.29)$$

where m is the current-density ratio. This voltage is PTAT, and has the interesting property that it is independent of I_{bias} . The resulting bias current can be found by combining (3.28) and (3.29):

$$I_{bias} = \frac{kT}{q R_{bias}} \ln m. \quad (3.30)$$

Such a current will be referred to as a PTAT/R current in this work. It only spreads as a result of the spread of R_{bias} .

Alternatively, as shown in Figure 3.12b, a single bipolar transistor can be used to generate a voltage that is complementary to absolute temperature (CTAT):

$$V_{bias} = V_{BE,bias} = \frac{kT}{q} \ln \left(\frac{I_{bias}}{I_S} \right). \quad (3.31)$$

The resulting bias current is CTAT/R, and is given implicitly by

$$I_{bias} = \frac{kT}{q R_{bias}} \ln \left(\frac{I_{bias}}{I_S} \right). \quad (3.32)$$

This current spreads as a result of spread of both I_S and R_{bias} , and will therefore be less reproducible than a PTAT/R current.

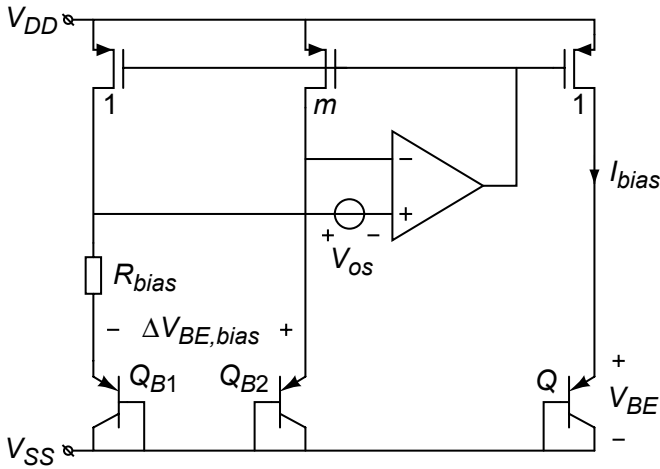


Figure 3.13. PTAT/R bias circuit.

Bias currents with other temperature dependencies can be generated by combining PTAT/R and CTAT/R currents. An example is a current proportional to a temperature-independent voltage (TI/R). This is needed for some curvature-correction techniques (see Section 3.5.3).

3.3.2 PTAT/R Bias Circuit

Figure 3.13 shows a practical circuit that can be used to generate a PTAT/R bias current [9]. Two diode-connected substrate pnp transistors Q_{B1} and Q_{B2} are biased at a $1 : m$ current ratio using two PMOS current sources. A feedback loop around an opamp ensures that the resulting difference in base-emitter voltage $\Delta V_{BE,bias}$ is generated across a resistor R_{bias} . As a result, the currents in the circuit will be PTAT/R currents, as described by (3.30). A third PMOS transistor is used to bias transistor Q , which provides the base-emitter voltage V_{BE} to be used for temperature sensing. This voltage then equals, ignoring non-idealities for now:

$$V_{BE,ideal} = \frac{kT}{q} \ln \left(\frac{\Delta V_{BE,bias}}{R_{bias} I_S} \right) = \frac{kT}{q} \ln \left(\frac{kT \ln m}{q R_{bias} I_S} \right). \quad (3.33)$$

As a result of various error sources in the circuit of Figure 3.13, the generated V_{BE} will deviate from this ideal value, which in turn will lead to temperature errors. The most important error sources are:

- offset V_{os} of the opamp,
- inaccuracy $\Delta m/m$ in the current mirror,

- finite open-loop gain A_{OL} .

The offset V_{os} adds directly to $\Delta V_{BE,bias}$ and thus introduces an error in the generated V_{BE} :

$$\begin{aligned} V_{BE} &= \frac{kT}{q} \ln \left(\frac{\Delta V_{BE,bias} + V_{os}}{R_{bias} I_S} \right) \\ &\simeq V_{BE}|_{V_{os}=0} + \frac{kT}{q} \frac{V_{os}}{\Delta V_{BE,bias}} = V_{BE}|_{V_{os}=0} + \frac{V_{os}}{\ln m}. \end{aligned} \quad (3.34)$$

Using (3.12), a maximum offset can be found given a maximum error contribution ΔT :

$$V_{os} < (3 \text{ mV} / ^\circ\text{C}) \cdot \ln m \cdot \Delta T. \quad (3.35)$$

Inaccuracy $\Delta m/m$ in the current mirror ratio modifies $\Delta V_{BE,bias}$:

$$\Delta V_{BE,bias} = \frac{kT}{q} \ln(m + \Delta m) \simeq \frac{kT}{q} \ln(m) + \frac{kT}{q} \frac{\Delta m}{m}. \quad (3.36)$$

The latter term adds to $\Delta V_{BE,bias}$ in the same way as V_{os} ; using (3.35) and substituting $T = 400 \text{ K}$ to get a worst-case value, the following requirement is obtained:

$$\frac{\Delta m}{m} < (9\% / ^\circ\text{C}) \cdot \ln m \cdot \Delta T. \quad (3.37)$$

Finite open-loop gain A_{OL} also results in an error: a finite overdrive voltage is present at the input of the amplifier which adds to $\Delta V_{BE,bias}$ and amounts to approximately $\Delta V_{BE,bias}/A_{OL}$. Using (3.35) again, and $T = 400 \text{ K}$, leads to the requirement

$$A_{OL} > \frac{\Delta V_{BE,bias}}{3 \cdot \ln m \cdot \Delta T} \simeq \frac{12 \text{ K}}{\Delta T}. \quad (3.38)$$

Note that A_{OL} is not exactly equal to the open-loop gain of the opamp, since the common-source PMOS transistors and R_{bias} are also in the loop.

Assume, for example, that $m = 10$, and that a maximum error contribution ΔT of 0.01 K is desired. The maximum offset is then $70 \mu\text{V}$, the maximum inaccuracy in the current mirror is 0.2% , and the open loop gain must be larger than 1200 , or 62 dB . An offset below $70 \mu\text{V}$ cannot be obtained using precision layout alone. A PTAT/R bias circuit in which the error due to offset is reduced by means of chopping will be presented in Section 7.3.5.

3.3.3 Compensation for Processing Spread

From Figure 3.5, it is clear that PTAT spread of V_{BE} , due to spread of I_S and R_{bias} , forms the main limitation to the untrimmed accuracy of CMOS smart

temperature sensors. While trimming can be used to correct for this spread, it increases the production costs. Therefore, it is worthwhile to investigate whether it is possible to generate a bias current that leads to a smaller PTAT spread by exploiting correlations between the spread of device parameters.

Several such compensation techniques have been proposed in literature:

- Correlation between I_S and the forward current-gain β_F : both parameters depend on the Gummel number. A high Gummel number corresponds to a low I_S and hence a high V_{BE} , and also to a low β_F and hence a high base current [10]. The voltage drop across an appropriately sized resistor in series with the base of Q_{B1} in Figure 3.13, will reduce the bias current for high Gummel numbers, and thus compensate to some extent for the higher value of V_{BE} [11].
- Correlation between I_S and a (pinched) well resistor: since the base of a substrate pnp is made of the same material as a (pinched) n-well resistor, correlation between the Gummel number and the resistance of well resistors can be expected (see Section 2.5.1). A high well doping results in a high Gummel number and hence a high V_{BE} , and at the same time in a small well-resistance [10]. The straightforward use of a well resistor as bias resistor therefore does not help: the bias current increases for high Gummel numbers, increasing V_{BE} even further [12]. If a current generated using a well resistor is subtracted from a current generated using a different resistor type, a bias current can be obtained that decreases for high Gummel numbers, so that some compensation for the higher V_{BE} may be obtained [13, 14].
- Correlation between I_S and the reverse current-gain β_R : for a relatively old bipolar process, a strong positive correlation between these two parameters has been reported [10]. In [15], it was proposed to exploit this correlation by generating a bias current proportional to β_R^2 . In CMOS, however, it is hard to use such a compensation circuit, because it requires operating a substrate pnp in reverse active mode, i.e. forward-biasing the n-well-substrate junction.

Unfortunately, little data on the actual *measured* performance of these compensation techniques is available in the open literature. This requires measurements of a large number of fabrication batches, or a measurements of a so-called ‘matrix-lot’, in which the wafers of a batch are split up in groups, each of which is exposed to different doping levels, corresponding to the corners of the process.

Some information could be obtained from measurements of the process control modules that are included on every wafer to monitor device parameters. However, if the base-emitter voltage of a substrate pnp transistor in such a module is measured along with, for instance, the transistor’s current gain or the n-well sheet resistance, the wafer temperature is usually not stabilized

better than $\pm 1^\circ\text{C}$ (which corresponds to $\pm 2\text{ mV}$ of variation of V_{BE}), so that conclusions are hard to draw.

Even if a compensation circuit can be used to reduce V_{BE} spread, spread of the bias resistor is likely to limit the best-case untrimmed inaccuracy. If an inaccuracy of $\pm 0.1^\circ\text{C}$ is desired, that spread has to be well below $\pm 1\%$, an unlikely low tolerance. Trimming, therefore, even with process compensation circuits, remains necessary for this level of accuracy.

3.4 Trimming

As discussed in the previous chapter, processing spread and packaging stress result in significant errors in the base-emitter voltage. In this section, trimming techniques are discussed that can be used to correct for these errors. A short overview of non-volatile memory techniques, needed to permanently store the trim setting, is also given.

3.4.1 Calculating Trimming Parameters

In order to trim a temperature sensor, its temperature error first needs to be established. This procedure, which will be referred to as ‘calibration’, is the topic of Chapter 6. Since calibration at more than one temperature is very expensive, a typical calibration procedure only provides information about the temperature error at one calibration temperature T_{cal} . There are many ways of trimming a sensor, i.e., many knobs that can be turned, so that the error *at* the calibration temperature is nulled. A sensor, however, has to meet a certain accuracy specification over a wider temperature range. Therefore, a trimming circuit has to exploit available knowledge about the temperature dependency of the error.

From the previous chapter, it is clear that errors in V_{BE} due to processing spread and mechanical stress are the dominant error sources, assuming that errors in the readout circuitry are negligible. While the temperature dependency of stress-induced errors is hard to predict accurately, processing spread mainly results in PTAT error in V_{BE} . Therefore, it makes sense to make a PTAT correction to V_{BE} :

$$V_{BE,trim} = V_{BE} + \gamma \frac{kT}{q}, \quad (3.39)$$

where the correction is expressed as a multiple of the thermal voltage kT/q .

The coefficient γ has to be calculated from calibration data. These data are: the reading D_{out} of the sensor before trimming and the actual temperature T_{cal} . The trimming parameter γ has to be chosen such that when V_{BE} is replaced by $V_{BE,trim}$, the reading of the sensor corresponds to the actual temperature.

Using equation (3.7), T_{cal} and D_{out} can be calculated back to an ideal ADC output μ_{ideal} and the actual output μ , respectively. Using (3.6), the error in

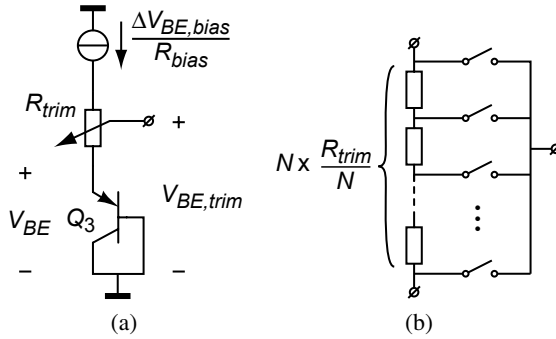


Figure 3.14. (a) Voltage-domain trimming by means of a programmable resistor in series with the emitter; (b) implementation of the programmable resistor.

V_{BE} can then be found:

$$V_{BE} - V_{BE,ideal} = \alpha \Delta V_{BE} \left(\frac{1}{\mu} - \frac{1}{\mu_{ideal}} \right). \quad (3.40)$$

This error should be cancelled by the PTAT voltage $\gamma \cdot kT/q$. Substituting (3.1) for ΔV_{BE} , we find

$$\gamma = -\alpha \ln(pr) \left(\frac{1}{\mu} - \frac{1}{\mu_{ideal}} \right). \quad (3.41)$$

Since all parameters in this equation are known, γ can be calculated.

The range of values that γ should be able to cover depends on the initial spread of V_{BE} . Using an estimate of ± 10 mV at 300 K, γ lies in the range $\pm 40\%$. The resolution with which γ can be set determines how accurately $V_{BE,trim}$ can be made equal to $V_{BE,ideal}$, and will thus limit the temperature error after trimming. Using (3.12), a maximum temperature error ΔT_{max} can be expressed as a minimum trimming resolution:

$$|V_{BE,trim} - V_{BE,ideal}| < (3 \text{ mV} / ^\circ\text{C}) \cdot \Delta T_{max}. \quad (3.42)$$

For example, a maximum temperature error $\Delta T_{max} = \pm 0.01$ $^\circ\text{C}$ corresponds to a maximum error in $V_{BE,trim}$ of ± 0.03 mV, and a maximum error in γ of $\pm 0.1\%$.

3.4.2 Trimming Circuits

Voltage-Domain Trimming

A straightforward way of making a PTAT correction to V_{BE} , in accordance with equation (3.39), is to literally add a programmable PTAT voltage to V_{BE} . Such a voltage can be a difference in base-emitter voltage ΔV_{BE} that is scaled by an appropriate factor γ_V .

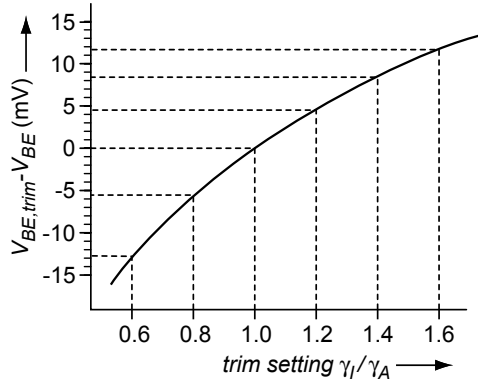


Figure 3.15. Logarithmic relation between a trim setting and the resulting adjustment in the base-emitter voltage for current-domain trimming (values are for $T = 300$ K).

One way of implementing this is shown in Figure 3.14a. A diode-connected substrate pnp transistor is biased at a PTAT/R current (generated, for instance, by the circuit of Figure 3.13). A resistor R_{trim} , of the same material as the resistor R_{bias} in the bias circuit, is added in series with the emitter of the transistor. The voltage drop across this resistor is therefore a PTAT voltage:

$$V_{BE,trim} = V_{BE} + \gamma_V \cdot \Delta V_{BE,bias} = V_{BE} + \gamma_V \cdot \frac{kT}{q} \ln m, \quad (3.43)$$

where $\gamma_V = R_{trim}/R_{bias}$ and m is the current-density ratio used for generating $\Delta V_{BE,bias}$. By adjusting the size of R_{trim} , the magnitude of the PTAT voltage can be adjusted. This can be done either continuously by means of laser trimming, or in discrete steps by dividing R_{trim} in N unit elements and selecting an appropriate tap using a multiplexer (Figure 3.14b). The latter option, however, requires a large number of elements for the trimming resolution required in a precision sensor.

Current-Domain Trimming

An alternative way of trimming V_{BE} is to adjust the emitter-current density $J_E = I_{bias}/A_E$. This can be achieved by changing I_{bias} by a factor γ_I and/or by changing the emitter area A_E by a factor γ_A :

$$V_{BE,trim} = \frac{kT}{q} \ln \frac{I_{bias}\gamma_I}{J_S A_E \gamma_A} = V_{BE} + \frac{kT}{q} \ln \frac{\gamma_I}{\gamma_A}, \quad (3.44)$$

where J_S is the saturation-current density.

With such current-domain trimming, the PTAT correction voltage is a logarithmic function of the trimming parameters γ_I and γ_A . This has to be taken into

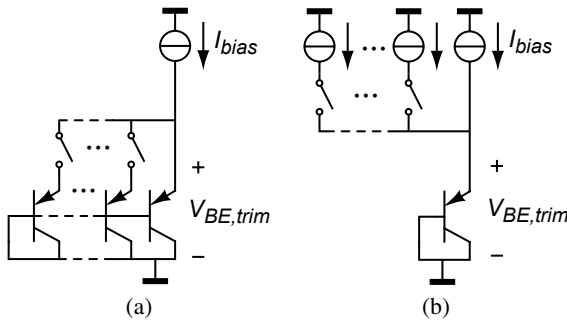


Figure 3.16. Current-domain trimming: (a) by means of a programmable emitter area; (b) by means of a programmable bias current.

account when determining the required range and resolution of these parameters. As illustrated in Figure 3.15, a given resolution of the trimming parameter corresponds to a smaller resolution of the PTAT correction voltage at the lower end of the trimming range.

The emitter area A_E can be made programmable by constructing the transistor from a parallel combination of sub-transistors with (binary-)scaled emitter areas, which can be switched on or off using a switch in series with the emitter (Figure 3.16a). A problem of this approach is that the voltage drop across the switches adds to the base-emitter voltages of the sub-transistors. Large switches may be required to make their on-resistance small enough for this voltage drop to be negligible.

The bias current I_{bias} can be programmed by switching a number of (binary-)weighted current sources (Figure 3.16b). Since the switches are now in series with the high-impedance output of current sources, their on-resistance can be large and small switches can be used. Trimming of the bias current may therefore be preferable over trimming of the emitter area.

Modulated Trimming

If a high trimming resolution is required, the previous trimming techniques require significant chip area as a result of the large number of switchable elements. An alternative that requires much less chip area is shown in Figure 3.17 [16]. The bias current of a diode-connected substrate pnp transistor is switched between two values I_{bias1} and I_{bias2} using a digital modulator. These values correspond to the extremes of the trimming range. (Alternatively, the switching can also be performed in the voltage domain, using a structure like that shown in Figure 3.14, but with only two taps.)

The modulator, which can be, for instance, a duty-cycle modulator or a digital sigma-delta modulator, produces an output signal whose time average can be programmed via a trim input γ_M ($0 \leq \gamma_M \leq 1$). The resulting base-emitter

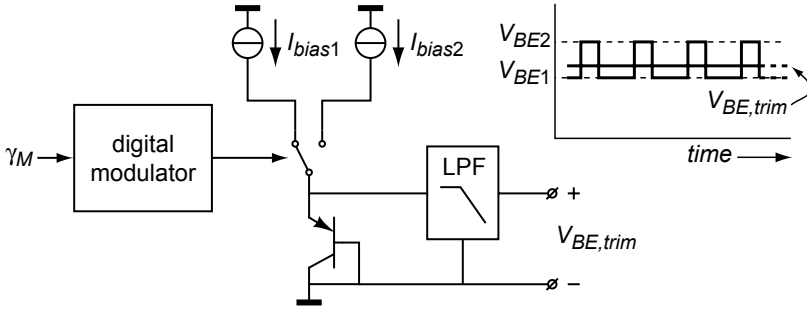


Figure 3.17. Trimming by programmable modulation of the bias current and averaging of the resulting switched base-emitter voltage.

voltage switches between two values V_{BE1} and V_{BE2} . A low-pass filter (LPF) is used to filter out the switching components, leaving the average value

$$V_{BE,trim} = (1 - \gamma_M) V_{BE1} + \gamma_M V_{BE2} = V_{BE1} + \gamma_M \cdot \frac{kT}{q} \ln \left(\frac{I_{bias2}}{I_{bias1}} \right). \quad (3.45)$$

Thus, by programming γ_M , the base-emitter voltage can be trimmed between V_{BE1} and V_{BE2} .

The area and complexity of the digital modulator can be kept to a minimum. An n -bit digital first-order sigma-delta modulator, for instance, can be realized using only an n -bit full-adder and an n -bit register [17]. A dedicated LPF may not be required: its function can be conveniently realized by the decimation filter of an oversampling ADC, as will be shown in Section 4.6.

Digital Trimming

Rather than trimming V_{BE} directly, it is also possible to modify the digital signal processing in the sensor so as to cancel the effect of V_{BE} spread, i.e. to trim the digital function (3.7) that converts the ADC's output to a temperature reading. The advantage of this approach is that the analog circuitry can be kept simple.

To evaluate what kind of adjustment has to be made to the digital processing, the effect of a PTAT error in V_{BE} on the ADC's output μ has to be determined. Substituting $V_{BE} = V_{BE,ideal} + \varepsilon \cdot kT/q$ in (3.6), the output μ can be written as

$$\mu = \frac{\alpha \cdot \Delta V_{BE}}{V_{BE,ideal} + \varepsilon \frac{kT}{q} + \alpha \cdot \Delta V_{BE}}, \quad (3.46)$$

while the ideal output is

$$\mu_{ideal} = \frac{\alpha \cdot \Delta V_{BE}}{V_{BE,ideal} + \alpha \cdot \Delta V_{BE}}. \quad (3.47)$$

The difference between the reciprocals of these outputs is a temperature-independent variable γ_D :

$$\frac{1}{\mu} - \frac{1}{\mu_{ideal}} = \frac{\varepsilon \frac{kT}{q}}{\alpha \cdot \Delta V_{BE}} = \gamma_D, \quad (3.48)$$

From this equation, μ_{ideal} can be expressed as a function of μ :

$$\mu_{ideal} = \frac{\mu}{1 - \gamma_D \mu}. \quad (3.49)$$

So ideally, the digital scaling expressed by equation (3.7) should be replaced by

$$D_{out} = A \cdot \frac{\mu}{1 - \gamma_D \mu} + B, \quad (3.50)$$

where A and B are constants, and γ_D is the trimming parameter.

The implementation of this non-linear function may require a substantial amount of chip area. It could be implemented in an external microcontroller, but this requires that the user of the sensor knows the trimming parameter and programs it into the microcontroller. A more attractive alternative is to use a least-squares linear fit to (3.50). Thus, the digital processing remains restricted to simple linear scaling, as in equation (3.7). Trimming is then achieved by adjusting the coefficients A and B . Thus, a fairly good performance can be obtained, as shown in Figure 3.18. A pure offset trim is even simpler to implement, but comes at the expense of larger errors towards the ends of the temperature range.

3.4.3 Trimming after Packaging

As described in Section 2.6, the change in V_{BE} due to a given level of mechanical stress is not PTAT (see Figure 2.11), as a result of the temperature dependency of the piezo-junction effect. Moreover, the applied stress itself is temperature dependent, as it results from the difference in thermal expansion coefficients between silicon and the packaging material. Typically, it will be zero at the temperature T_p at which the die and packaging material have been joined together. For plastic packages, this is the molding temperature of 175 °C. The further away from this temperature, the larger the stress will be [18]. These two effects result in spread of V_{BE} that is certainly not PTAT. This spread will be especially significant in plastic packages. It is therefore important to evaluate what errors remain if it is corrected for by means of PTAT trimming.

The exact temperature dependency of packaging-induced spread is hard to predict, as it depends on many factors (it can even be unstable during thermal

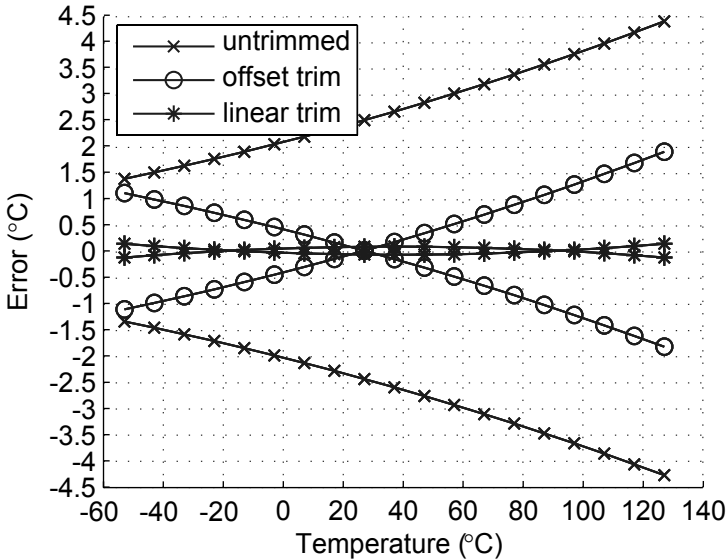


Figure 3.18. Temperature errors resulting from a PTAT error in V_{BE} of ± 10 mV at 300 K: untrimmed, trimmed using digital offset correction, and trimmed using a linear correction (offset and gain).

cycling). Figure 3.19a gives a qualitative picture of the total spread of V_{BE} as a result of the combined effects of processing spread and packaging. If a PTAT correction is used to trim V_{BE} after packaging, the processing-related PTAT spread is eliminated. As shown in Figure 3.19a, the packaging-induced spread, in contrast, is only nulled at the temperature T_r at which V_{BE} is trimmed to its nominal value. Due to the difference in temperature dependency between the packaging-induced spread and the PTAT correction, some spread remains which increases towards the low and high ends of the temperature range. The resulting temperature errors will have the same characteristic: they will be small around the calibration temperature and increase towards the extremes of the temperature range.

The difference in temperature dependency between spread induced by packaging and processing-related spread makes it difficult to obtain high accuracy over the full operating temperature range by means of calibration at a single temperature only. Adding a second temperature is usually not a cost-effective solution. A feasible alternative is to combine wafer-level trimming and post-packaging trimming. On the wafer, the stress-induced changes in V_{BE} are negligible [18]. PTAT trimming can therefore be used to accurately correct for process-related spread. The spread after packaging is then purely a result

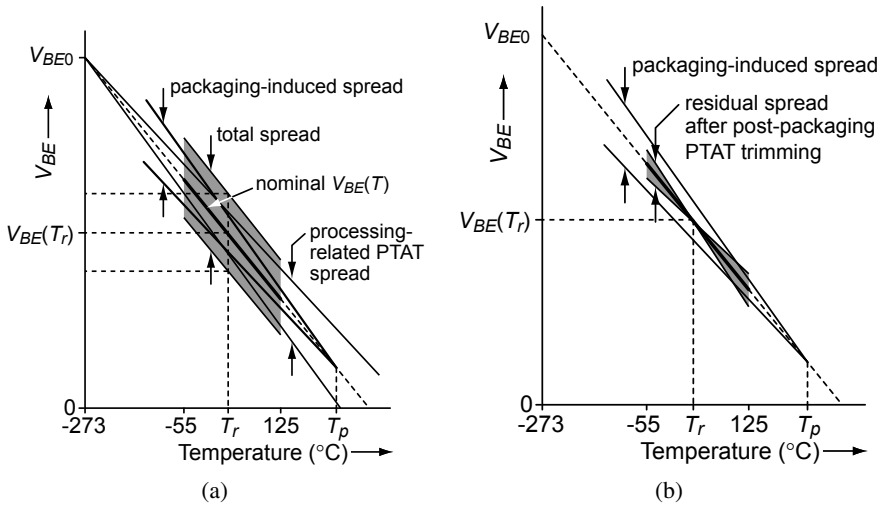


Figure 3.19. (a) Total spread of the base-emitter voltage after packaging, consisting of processing-related spread, and packaging-induced spread; (b) spread remaining after post-packaging trimming.

of packaging stress, and can be trimmed using a correction voltage that better mimics the temperature dependency of the stress-induced changes.

The packaging-induced spread shown in Figure 3.19a is symmetrical around the nominal value of V_{BE} . In practice, this is however not always the case. Certainly for plastic packages, which expose the die to compressive stress, there will be a systematic negative shift in V_{BE} , plus a random spread around that shift [18, 19]. If this systematic shift is significant compared to the processing-related spread, PTAT trimming after packaging is not advisable. In that case, it is probably best to eliminate processing-related spread by means of wafer-level PTAT trimming. The packaging shift can then be corrected for either using a systematic correction, or by means of a second trimming step after packaging.

3.4.4 Non-Volatile Memory Technology

Irrespective of what trimming technique is used, a permanent adjustment to the chip needs to be made during production. In precision analog chips (e.g. low-offset amplifiers or bandgap references), such an adjustment is usually made by means of laser trimming. The value of a resistor is then increased by making cuts in it using a laser beam. Tolerances down to 0.01% can be achieved [20], though stability over time and temperature may be a problem. Laser trimming is obviously not possible after packaging. Trimming of polysilicon

resistors after packaging can be achieved by passing current pulses through them. The local heating associated with these pulses results in a permanent decrease of the resistor value [21].

In precision CMOS analog circuits, such as CMOS temperature sensors, trimming is usually done in discrete steps, as described in the previous sections. In that case, a digital non-volatile memory is used that drives either analog switches or digital correction circuitry. Such digital non-volatile memory can either be erasable or non-erasable. Non-erasable memory is sometimes referred to as OTP (One-Time Programmable). OTP is sufficient for temperature sensors, but CMOS foundries sometimes only offer erasable memory (e.g. EPROM) as a standard extension to the baseline CMOS process. Therefore, both types will be discussed briefly.

The two most common non-erasable digital non-volatile memory techniques are zener zapping and fusible links:

- *Zener zapping* changes a zener diode, which initially acts as an open-circuit, to a short circuit. This is done by bringing it in avalanche mode, which destroys the junction and creates a reliable metallic connection [22]. Relatively high programming voltages ($> 6\text{ V}$) are required to bring the zener in avalanche mode. These voltages have to be processed with special care to avoid breakdown of other junctions. In CMOS, a lateral zener diode can be constructed from adjacent n^+ and p^+ diffusions [23].
- *Fusible links* consist of a metal or polysilicon connection that can be physically destroyed by passing a large current through it [24]. An initial short-circuit is thus converted to an open-circuit. An advantage compared to zener zapping is that low-voltage pulses can be used ($< 7\text{ V}$), which are below the junction breakdown voltage of most CMOS devices. Fusible links may, however, be less reliable than zener zapping, because metal regrowth can (partially) restore the connection. Since this recovery is never complete, the reliability can be increased by comparing the resistance with that of an unblown fuse [25]. Links can also be broken by cutting them with a laser beam [24]. This simplifies the circuitry, but again has the disadvantage that it cannot be performed after packaging.

Most *erasable* non-volatile memories are based on floating-gate technology [26]: the threshold of a MOS transistor is altered by storing charge on an extra floating polysilicon gate in between the selection gate of the transistor and its channel. In the case of EPROM (Electrically Programmable Read-Only Memory), programming consists of injection of hot-electrons from the drain into the floating gate as a result of a high voltage ($> 10\text{ V}$) being applied to the selection gate. This charge can be released by exposing the chip to ultraviolet light. In the case of EEPROM (Electrically Erasable EPROM), the charge on the floating gate can be removed electrically by means of tunneling through

a thin oxide layer. Floating-gate technology is mostly used to create digital memory, but analog (continuous) trimming has also been reported [27].

3.5 Curvature Correction

Curvature of the base-emitter voltage is mainly an issue in bandgap references used in general-purpose measurement systems, where a temperature-dependent reference introduces cross-sensitivity to temperature. In a smart temperature sensor, in contrast, temperature is the measurand, and curvature therefore only results in non-linearity. As a result, in addition to the many curvature-correction techniques that have been developed for bandgap references, linearization techniques tailored to curvature-correction in smart temperature sensors have also been developed.

This section starts with an analysis of the temperature errors due to curvature. Curvature in smart temperature sensors will be compared to that in bandgap voltage references, and an overview of curvature-correction techniques for bandgap voltage references will be given. In addition, some curvature correction techniques will be described that are specifically intended for temperature sensors.

3.5.1 Errors due to Curvature

As derived in Section 2.3, the temperature dependency of the base-emitter voltage of a bipolar transistor can be described by equation (2.33):

$$V_{BE}(T) = \underbrace{V_{BE0} - \lambda T}_{\text{tangent at } T=T_r} + c(T), \quad (3.51)$$

where the curvature $c(T)$ is given by equation (2.39),

$$c(T) = \frac{k}{q} (\eta - m) \left(T - T_r - T \ln \left(\frac{T}{T_r} \right) \right), \quad (3.52)$$

if the bias current is assumed to be proportional to T^m (see Figure 2.4b).

Substituting these equations in the transfer (3.6) of a smart temperature sensor, we find:

$$\begin{aligned} \mu &= \frac{V_{PTAT}}{V_{REF}} = \frac{\alpha \cdot \Delta V_{BE}}{V_{BE} + \alpha \cdot \Delta V_{BE}} \\ &= \frac{\frac{kT}{q} \ln(pr)}{V_{BE0} - \lambda T + \frac{kT}{q} \ln(pr) + c(T)}. \end{aligned} \quad (3.53)$$

where (3.1) has been substituted for ΔV_{BE} . If the denominator V_{REF} is a conventional bandgap reference voltage, the linear term in $V_{BE}(T)$ is cancelled by $\alpha \cdot \Delta V_{BE}$ so as to obtain a zero temperature coefficient at $T = T_r$:

$$\left[\frac{\partial V_{REF}}{\partial T} \right]_{T=T_r} = 0, \quad (3.54)$$

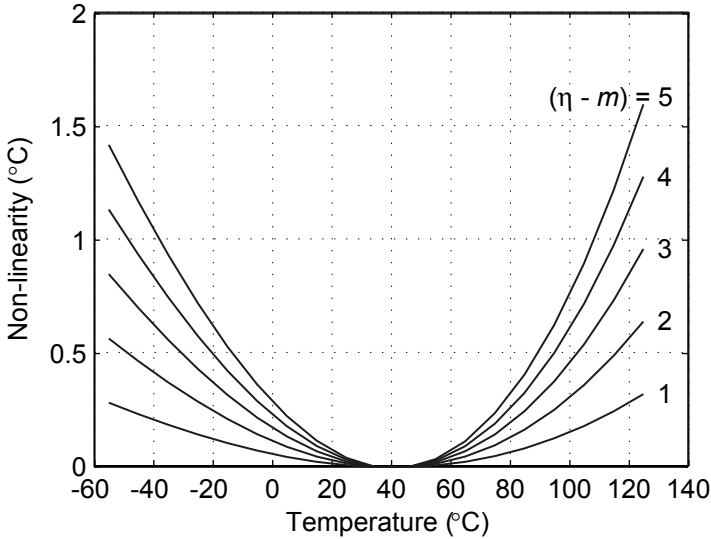


Figure 3.20. Non-linear temperature error resulting from the curvature of V_{BE} , for various values of $(\eta - m)$.

which implies

$$\lambda = \frac{k}{q} \ln(pr) \quad (3.55)$$

and

$$V_{REF}(T_r) = V_{BE0}. \quad (3.56)$$

The curvature of V_{BE} then results in an equal curvature in V_{REF} , and hence in an inverse non-linearity in μ .

If the digital output D_{out} is obtained from μ by simple linear scaling, this non-linearity results in a systematic temperature error, which is shown in Figure 3.20 for various values of $(\eta - m)$. A typical value of $\eta - m$ is 3, since CMOS substrate pnp transistors have a typically value of $\eta \simeq 4$ [28], and $m = 1$ for a PTAT/R bias current generated using a temperature-independent resistor. The associated temperature error of almost 1 °C is unacceptably large for precision applications. Some form of curvature correction is therefore needed.

3.5.2 Comparison to Voltage References

Many curvature corrections techniques have been presented in literature that aim at minimizing the temperature dependency of the output of general-purpose bandgap voltage references. In principle, these techniques can also be applied in smart temperature sensors. The effectiveness of curvature correction in a

voltage reference is usually expressed in terms of its residual temperature coefficient (TC), which is often calculated using the so-called box-method [29]. The TC α_{TCREF} then equals the relative peak-to-peak variation of the reference voltage over a given temperature range divided by the width of that range, i.e.:

$$\alpha_{TCREF} = \frac{V_{REF,max} - V_{REF,min}}{V_{REF,ideal}(T_{max} - T_{min})}. \quad (3.57)$$

For comparison purposes, it is useful to translate this TC to an equivalent temperature error in a smart temperature sensor. If the reference voltage in a temperature sensor has a given residual temperature coefficient, the maximum variation in V_{REF} over the operating temperature range is given by:

$$|V_{REF} - V_{REF,ideal}| < \frac{1}{2} \cdot (T_{max} - T_{min}) \cdot V_{REF,ideal} \cdot \alpha_{TCREF}, \quad (3.58)$$

where it is assumed that the ideal reference voltage $V_{REF,ideal}$ is halfway between $V_{REF,min}$ and $V_{REF,max}$. The temperature error that corresponds to this variation can be found by multiplying with the sensitivity of the digital output D_{out} to variations in V_{REF} , which is $-T/V_{REF}$. This gives a maximum temperature error of

$$\begin{aligned} |\Delta T| &= S_{V_{REF}}^{D_{out}} \cdot |V_{REF} - V_{REF,ideal}| \\ &< \frac{1}{2} \cdot (T_{max} - T_{min}) \cdot T_{max} \cdot \alpha_{TCREF}. \end{aligned} \quad (3.59)$$

Over the military temperature range of -55°C to 125°C , this boils down to $3.6 \cdot 10^4 \text{ K}^2 \cdot \alpha_{TCREF}$. A TC of $10 \cdot 10^{-6} \text{ K}^{-1}$ for instance, leads to maximum temperature error of 0.36 K.

Conversely, if a maximum temperature error ΔT is desired, the TC of the reference should be smaller than

$$\alpha_{TCREF} < (28 \cdot 10^{-6} \text{ K}^{-2}) \cdot \Delta T. \quad (3.60)$$

For maximum temperature of $\pm 0.1^\circ\text{C}$, a TC of less than $\pm 2.8 \cdot 10^{-6} \text{ K}^{-1}$ is required. This is a challenging number, given that the highest-performance commercial voltage references (based on zener diodes) are not much better, having TCs in the order of $1 \cdot 10^{-6} \text{ K}^{-1}$ [30,31]. The TCs of bandgap references reported in literature are often much higher. It is however important to realize that these higher numbers often can be attributed to other effects than curvature, such as the effects offset and mismatch.

3.5.3 Curvature-Correction Techniques for Bandgap Voltage References

The many published curvature-corrected bandgap voltage references can be classified as follows:

- references in which V_{BE} is linearized using a temperature-dependent bias current;
- references in which a temperature-dependent gain is used for adding ΔV_{BE} to V_{BE} ;
- references in which a compensating non-linearity is generated in ΔV_{BE} using a temperature-dependent bias current ratio;
- references in which a piecewise-linear voltage is used to compensate for the curvature of V_{BE} ;
- references in which a compensating non-linearity is generated some other way.

In the following, these categories will be explained in more detail. At the end of the section, the techniques will be compared in terms of their suitability for use in temperature sensors.

Using a Temperature-Dependent Bias Current

Since the curvature of V_{BE} is a result of the strong non-linear temperature dependency of the saturation current I_S , it can be compensated for using a bias current that is also strongly temperature-dependent. From equation (3.52), it can be seen that a bias current proportional to T^m leads to a curvature that is proportional to $(\eta - m)$. This explains why a PTAT/R bias current ($m = 1$) results in a smaller curvature than a TI/R bias current ($m = 0$) [32].

If a bias current could be generated for which $m \simeq \eta$, the majority of the curvature could be cancelled. This idea was mentioned in an early paper of Song and Gray [33], and developed by Filanovsky, who used a translinear circuit to generate a bias current that is a linear combination of a PTAT³/R and a PTAT⁴/R current, corresponding to typical values of η between 3 and 4 [34]. The simulated TC of his 0.8 μm BiCMOS implementation was $8 \cdot 10^{-6} \text{ K}^{-1}$. This technique is however hard to implement in pure CMOS, because the fairly complex translinear circuit relies heavily on the availability of free bipolar transistors.

As discussed in Section 2.8.2, the temperature dependency of the bias current, and hence also the curvature, are affected by the temperature dependency of the bias resistor R_{bias} . Resistors with a negative first-order temperature coefficient (α_{TCR1}) reduce the curvature, while (practical) resistors with a positive α_{TCR1} increase the curvature. It can be shown that theoretical optimal values for α_{TCR1} exist for which the second-order curvature is nulled [35, 36]:

$$\alpha_{TCR1} = \frac{1 \pm \sqrt{1 + \eta - m}}{T_r}. \quad (3.61)$$

For typical values $\eta = 4$, $m = 1$, and $T_r = 300$ K, the optimal coefficients are $-3.3 \cdot 10^{-3} \text{ K}^{-1}$ and $10 \cdot 10^{-3} \text{ K}^{-1}$. Unfortunately, as can be seen in Table 2.2, resistors with such high TCs are usually not available in standard CMOS processes. High-resistivity poly resistors come closest.

Falconi *et al.* proposed to combine two resistors of different type in a so-called ‘anti-series’ or ‘anti-shunt’ combination, so as to create an effective bias resistor with a large TC [36]. An important disadvantage of this approach is that the ratio of these two resistors cannot be made accurate by design and has to be trimmed. Moreover, the effective bias resistor will not match with other resistors, as is often required.

Using a Temperature-Dependent Gain

If the gain α applied to ΔV_{BE} is made slightly temperature-dependent, a non-linearity can be introduced that compensates for the curvature of V_{BE} . As the gain α often depends on a resistor ratio, a temperature-dependent α can be obtained by using resistors with different TCs. This idea was introduced by Lewis and Brokaw, who used a series combination of a low-TC and a high-TC resistor in bipolar technology [37]. The idea was extended to higher-orders by Audy, who introduced more degrees of freedom by using a more complicated combination of resistors (still using two resistor types) [38, 39]. His bipolar implementations achieved TCs below $5 \cdot 10^{-6} \text{ K}^{-1}$. The idea was ported to CMOS by Leung *et al.*, who used a combination of high-ohmic poly resistors and p^+ active resistors [40]. Their implementation in $0.6 \mu\text{m}$ CMOS achieved (after calibration at multiple temperatures) a TC of about $6 \cdot 10^{-6} \text{ K}^{-1}$. A common disadvantage of all these implementations is that the ratio of the resistors has to be trimmed.

Using a Temperature-Dependent Bias Current Ratio

A compensation voltage that naturally matches the curvature of V_{BE} can be generated from the difference in base-emitter voltage between two transistors biased at a temperature-dependent current *ratio*. This idea was introduced in a seminal paper by Widlar [41] and has since then returned in literature in many different forms. It can be best illustrated by considering a transistor biased at a PTAT/R collector current,

$$I_{PTAT/R} = \frac{\lambda_{bias} T}{R_{bias}}, \quad (3.62)$$

and a transistor biased at a current derived from a temperature-independent (TI) voltage V_{TI} using a resistor (TI/R),

$$I_{TI/R} = \frac{V_{TI}}{R_{bias}}, \quad (3.63)$$

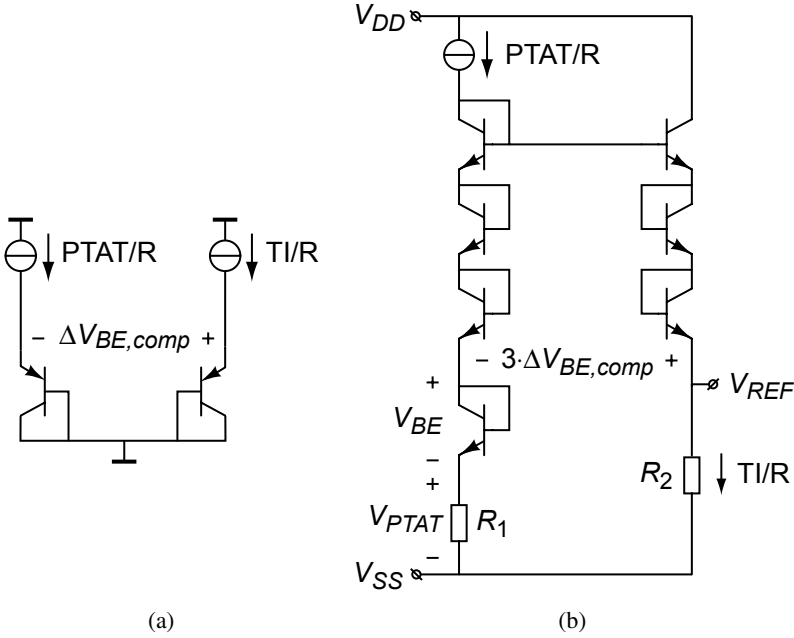


Figure 3.21. Curvature-correction using a temperature-dependent bias current ratio: (a) generation of the compensation voltage; (b) bipolar bandgap voltage reference that uses this technique [42].

where V_{TI} is obtained, for example, using feedback from the output of the curvature-corrected reference. The difference in base-emitter voltage between these transistors is then (Figure 3.21a):

$$\begin{aligned} \Delta V_{BE,comp} &= \frac{kT}{q} \ln \left(\frac{I_{PTAT/R}}{I_{TI/R}} \right) = \frac{kT}{q} \ln \left(\frac{\lambda_{bias} T}{V_{TI}} \right) \\ &= \frac{k}{q} \left(T \ln \left(\frac{T}{T_r} \right) + T \ln \left(\frac{\lambda_{bias} T_r}{V_{TI}} \right) \right). \end{aligned} \quad (3.64)$$

This has the same $T \ln(T)$ characteristic as found in the curvature of V_{BE} , equation (3.52). By adding an appropriately scaled $\Delta V_{BE,comp}$ to V_{BE} , the curvature can therefore be cancelled almost completely. The scale factor will roughly be $(\eta - 1)$; its optimal value is such that any additional curvature due to the temperature dependency of R_{bias} is also cancelled as much as possible. A small residual curvature remains, resulting from the fact that the R_{bias} -dependent term will not exactly fit the $T \ln(T)$ characteristic.

Meijer *et al.* presented a bipolar implementation of this technique in which the scaling of $\Delta V_{BE,comp}$ is realized by stacking bipolar transistors [42] (Figure 3.21b). The number of stacked transistors has to be chosen as close as possible to

$(\eta-1)$ (3 in Figure 3.21b). The difference results in some residual curvature, but this was very low for the reported implementation: after calibration at multiple temperatures, the TC could be trimmed to $0.5 \cdot 10^{-6} \text{ K}^{-1}$. The expected TC for a single-temperature calibration was $5 \cdot 10^{-6} \text{ K}^{-1}$. Obvious disadvantages of this implementation are the high supply voltage required and the difficulty of porting it to CMOS.

An implementation in $5 \mu\text{m}$ CMOS was presented by Lin and Salama [43]. They implemented the scaling of $\Delta V_{BE,comp}$ by means of a resistor ratio, so that non-integer values can be used. After a single-temperature trim, they obtained a typical TC of $15 \cdot 10^{-6} \text{ K}^{-1}$. This relatively high value is attributed to offset of the opamps used (in contrast with bipolar opamps, the offset of a CMOS opamp tends to have a non-PTAT characteristic and cannot be trimmed out completely).

Gunawan *et al.* presented a current-mode implementation of essentially the same technique in bipolar technology, which is capable of operating from a 1 V supply. They convert V_{BE} , ΔV_{BE} and $\Delta V_{BE,comp}$ into currents, which are then summed into a resistor with appropriate scaling factors to obtain a sub-bandgap curvature-corrected reference voltage. After a single-temperature trim, a TC of $4 \cdot 10^{-6} \text{ K}^{-1}$ was obtained.

A low-voltage current-mode implementation in CMOS of again essentially the same technique, presented by Malcovati *et al.* [44], is shown in Figure 3.22. Resistor R_1 converts ΔV_{BE} of transistors Q_1 and Q_2 to a current, while resistors R_2 carry currents proportional to V_{BE} . With appropriate scaling of the resistors, the sum of these currents is a TI/R current, but still with curvature. A third transistor Q_3 is biased by this TI/R current, and the difference between its base-emitter voltage and that of Q_1 , which is the compensation voltage $\Delta V_{BE,comp}$, is converted to currents using resistors R_4 . As a result, the total output current, when passed through a resistor R_3 , results in a curvature-compensated reference voltage. The BiCMOS prototype of Malcovati *et al.* operated from a 1 V supply and achieved a TC of $7.5 \cdot 10^{-6} \text{ K}^{-1}$. The BiCMOS technology was only needed for the low-voltage implementation of the opamp. The technique works just as well in CMOS technology, but a somewhat larger supply voltage may be needed for the opamp.

A switched-capacitor implementation of this technique was used in a CMOS temperature sensor by Hagleitner [45], who obtained a linearity better than $\pm 0.2^\circ\text{C}$.

Palmer and Dobkin used a slightly different approach [46]: in their implementation in bipolar technology, they created a compensating non-linearity in the main ΔV_{BE} of a bandgap reference, rather than using an additional $\Delta V_{BE,comp}$. They biased the transistors used for generating ΔV_{BE} at a CTAT/R current, but created a PTAT/R unbalance in the bias currents. The resulting

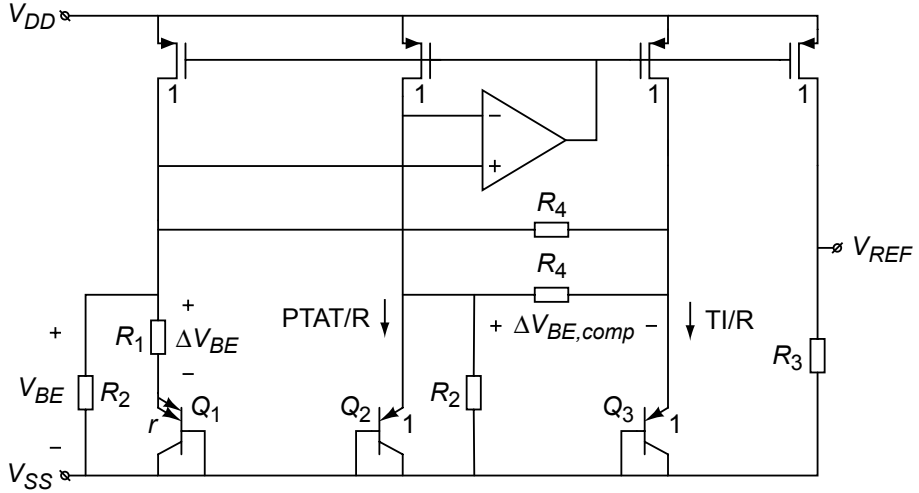


Figure 3.22. CMOS low-voltage bandgap voltage reference that uses a temperature-dependent bias current ratio for curvature-correction [44].

non-linearity in ΔV_{BE} compensates for the curvature of V_{BE} , leading to a TC of about $1 \cdot 10^{-6} \text{ K}^{-1}$ after trimming.

Song and Gray used essentially the same principle in CMOS [33]: two substrate pnp transistors with an $r : 1$ emitter-area ratio are bias at a TI/R current, while a sourcing and a sinking PTAT/R current source are used to create an unbalance in their bias currents (Figure 3.23). The resulting ΔV_{BE} is

$$\begin{aligned}
 \Delta V_{BE} &= \frac{kT}{q} \ln \left(r \frac{I_{TI/R} + I_{PTAT/R}}{I_{TI/R} - I_{PTAT/R}} \right) \\
 &= \frac{kT}{q} \ln(r) + 2 \frac{kT}{q} \left(\frac{I_{PTAT/R}}{I_{TI/R}} \right) \\
 &\quad + \frac{2}{3} \frac{kT}{q} \left(\frac{I_{PTAT/R}}{I_{TI/R}} \right)^3 + \dots
 \end{aligned} \tag{3.65}$$

If the PTAT/R current is zero, this is a conventional PTAT voltage. A non-zero PTAT/R unbalance results in an additive quadratic correction voltage that can be scaled so as to cancel the curvature of V_{BE} . The reported switched-capacitor implementation in $6 \mu\text{m}$ CMOS had a typical TC of $25.6 \cdot 10^{-6} \text{ K}^{-1}$ after trimming at room temperature.

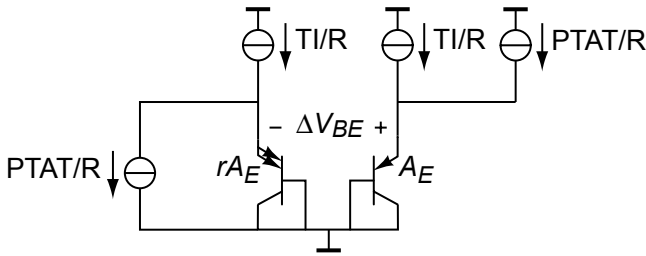


Figure 3.23. Generation of a curvature-correcting non-linearity in ΔV_{BE} using a PTAT/R unbalance in the bias currents [33].

Piecewise-Linear Correction

A further way of compensating the curvature of a bandgap voltage reference is to somehow generate a compensation voltage of which the temperature dependency mimics the curvature of V_{BE} . One possibility is to use a piecewise-linear correction voltage. The operating temperature range is then divided in a number of segments, in each of which the curvature is linearized. Comparison of appropriately scaled versions of V_{BE} and ΔV_{BE} can be used to detect in which segment the operating temperature lies, and to add compensation voltages accordingly. These compensation voltages are typically also composed of scaled versions of V_{BE} and ΔV_{BE} . Rincón-Mora described a reference, implemented in $2\ \mu\text{m}$ BiCMOS, which uses two segments [29]. The comparison and generation of the compensation have been implemented in the current domain. After calibration over the full operating range and trimming, a TC below $20 \cdot 10^{-6}\ \text{K}^{-1}$ was obtained. Bakker reported a similar CMOS implementation using 6 segments which achieves a TC below $10 \cdot 10^{-6}\ \text{K}^{-1}$ [2].

Other Correction Techniques

Lee *et al.* used the exponential temperature dependency of the forward current-gain β_F to generate a non-linear compensation voltage [47]. This voltage is generated by passing the sum of a PTAT/R current and the base current of a bipolar transistor biased at a PTAT/R collector current through a resistor. By appropriately scaling the resulting PTAT and PTAT/β_F components, curvature can be minimized. A disadvantage of this technique is that spread of β_F will increase the spread of the reference voltage. Only PTAT components of this spread can be removed with a single-temperature trim. For the $1.5\ \mu\text{m}$ BiCMOS process used, a TC below $10 \cdot 10^{-6}\ \text{K}^{-1}$ over the military temperature range was reported after trimming for minimum TC.

A non-linear correction voltage can also be generated using MOS transistors. Chin and Wu proposed to add the difference in gate-source voltage ΔV_{GS} of

two MOS transistors to ΔV_{BE} in a bandgap cell, so that its output voltage is the sum of a base-emitter voltage, an amplified ΔV_{BE} , and an amplified ΔV_{GS} [48]. The latter term has a non-linear temperature dependency that can be tuned by appropriate sizing of the MOS transistors to compensate for the curvature of V_{BE} . A typical TC of $5.5 \cdot 10^{-6} \text{ K}^{-1}$ was reported for an implementation in $3.5 \mu\text{m}$ CMOS, but it is unclear what type of trimming was used. With this type of correction, threshold-voltage mismatch will affect the accuracy of the reference voltage.

Salminen and Halonen proposed to add a diode-connected PMOS transistor in series with the bipolar transistor that generates V_{BE} [49]. The non-linear temperature dependency of the gate-source voltage that is thus added to the reference voltage is used to correct for curvature. While the technique looks promising in simulation, measurements show a fairly large spread and large temperature coefficients (about $100 \cdot 10^{-6} \text{ K}^{-1}$), which can be attributed to the fact that the full threshold voltage spread (albeit not amplified) affects the output voltage, rather than threshold-voltage mismatch.

Comparison

A comparison of the above techniques in terms of their performance is hard to make, since the cited implementations differ in many more respects than just the way curvature is corrected. For instance, the way a reference is trimmed (based on a single-temperature calibration or based on calibration at multiple temperatures) has a big influence on the TC that is obtained. Moreover, the degree at which other non-idealities than curvature (such as offset and mismatch) are addressed also affects the performance. A qualitative comparison, however, can be made.

The described techniques can be divided in two groups: those that can be made accurate by design, and those that require trimming. The first group includes the use of a temperature-dependent bias current ratio and piecewise-linear correction, while the latter group includes the techniques based on resistors with different TCs, the non-linear temperature dependency of the current gain, and the temperature dependency of the gate-source voltage of MOS transistors. These latter techniques require trimming because they are based on non-ideal device characteristics that are subject to processing spread.

The techniques that require trimming are obviously not to be preferred for untrimmed references or temperature sensors, as they will decrease their initial accuracy. But even for references and temperature sensors that will be trimmed, they are not desirable, because they introduce additional degrees of freedom in the spread of the reference voltage or the temperature reading, which may call for an expensive calibration at multiple temperatures.

Since curvature is considered to be a systematic error (i.e. with negligible spread), it calls for systematic correction that can be made accurate by design.

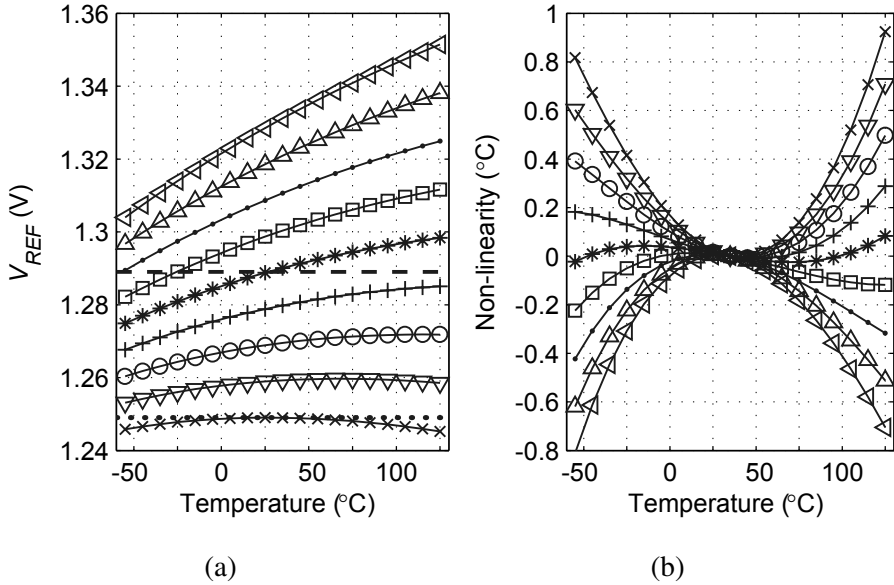


Figure 3.24. (a) Reference voltage V_{REF} , for various values of $V_{REF}(T_r)$ ($\eta - m = 3$); the dotted line indicates V_{BE0} , the dashed line the optimum value of $V_{REF}(T_r)$ given by equation (3.69); (b) the corresponding non-linearity in the ADC's output μ .

The use of a temperature-dependent bias current ratio based on a PTAT/R current and a TI/R current is such a technique. The same goes for piecewise-linear correction.

3.5.4 Ratiometric Curvature Correction

So far, it has been assumed that the reference voltage V_{REF} is a conventional first-order compensated bandgap voltage reference, i.e. that the temperature coefficient (TC) of V_{REF} is zero at T_r . Although this may appear a logical choice, it is not optimal in terms of the linearity of the ADC's output μ [50, 51]. This can be shown if V_{REF} and the non-linearity of μ are plotted for different values of $V_{REF}(T_r)$ (Figure 3.24). If $V_{REF}(T_r)$ equals V_{BE0} (dotted line), it has a zero TC at $T = T_r$, and the associated non-linearity is quadratic and amounts to almost 1°C . If $V_{REF}(T_r)$ is increased, its TC becomes slightly positive, and the non-linearity decreases to a third-order term of less than 0.2°C for $V_{REF}(T_r) \simeq 1.29\text{ V}$ (dashed line).

This reduced non-linearity can be explained as follows. A positive linear temperature dependence of V_{REF} results in a non-linearity in μ :

$$\mu \propto \frac{1}{V_{REF}} \propto \frac{1}{1 + xT} = 1 - xT + \frac{1}{2}(xT)^2 - \dots \quad (3.66)$$

This non-linearity can be tuned to compensate for the second-order non-linearity originating from the curvature of V_{BE} . What remains is a third-order non-linearity. As this type of curvature correction relies on the ratiometric nature of the temperature sensor, it will be referred to as *ratiometric curvature correction*.

The condition under which this minimum non-linearity is achieved can be found by requiring that

$$\left[\frac{\partial^2 \mu}{\partial T^2} \right]_{T=T_r} = 0, \quad (3.67)$$

Substitution of (3.53) and (3.51) gives, after some manipulation,

$$V_{REF}(T_r) = V_{BE}(T_r) + \alpha \cdot \Delta V_{BE}(T_r) = \frac{V_{BE0}}{V_{BE0} + \frac{1}{2} T_r^2 c''(T_r)} V_{BE0}. \quad (3.68)$$

If the curvature is given by equation (3.52), the expression for $V_{REF}(T_r)$ can be rewritten as

$$V_{REF}(T_r) = \frac{V_{BE0}}{V_{BE0} - \frac{k}{2q} (\eta - m) T_r} V_{BE0}, \quad (3.69)$$

which is, as expected, slightly larger than V_{BE0} . This level is indicated by the dashed line in Figure 3.24a. In practice, the value found using these equations will be close to the optimum, but some fine-tuning by means of simulation may be required, because the condition (3.67) leads to a local minimum around T_r , and not necessarily to the minimum peak-to-peak non-linearity in a larger temperature range.

Ratiometric curvature correction does not alter the structure of the temperature sensor (compared to that shown in Figure 3.1), but merely its parameters. Slightly different values for the scaling parameters A and B in equation (3.7) will be needed to obtain a digital output in degrees Celsius. In fact, a proper choice of these parameters ensures that a trimmed sensor will automatically have minimum curvature. This is because trimming essentially changes the value of $V_{BE}(T_r)$, and hence that of $V_{REF}(T_r)$, and with the right choice of A and B , a zero temperature error corresponds to a value of $V_{REF}(T_r)$ that satisfies equation (3.68).

By means of ratiometric curvature correction, the non-linearity is reduced significantly compared to the non-linearity that is obtained with a conventional bandgap reference. Further reduction, as required for inaccuracies in the order of $\pm 0.1^\circ\text{C}$, can be achieved, for instance, by a small additional correction in the digital domain (see Section 4.5.4), or by using higher-order ratiometric correction, as will be discussed in the next section.

3.5.5 Higher-Order Ratiometric Curvature Correction

The ratiometric curvature correction described in the previous section only eliminates the second-order non-linearity. As can be seen in Figure 3.24b, an

S-shaped third-order non-linearity remains. To be able to eliminate this non-linearity as well, extra degrees of freedom have to be created. This can be done by using *multiple* analog-to-digital conversions [52]. Each of these conversions results in a ratio μ_i , similar to the ratio μ that has been considered so far, but for each conversion different values for the parameters are used, e.g. for the gain α . The digital output D_{out} is a linear combination of these ratios:

$$D_{out} = \sum_i A_i \mu_i + B. \quad (3.70)$$

The coefficients A_i and the parameters of the individual ratios can now be chosen such that the non-linearity, including higher-orders, is minimized.

The following combination is a simple example of higher-order ratiometric curvature correction using two AD conversions:

$$D_{out} = A(\mu_1 - c \cdot \mu_2) + B, \quad (3.71)$$

$$\mu_1 = \frac{\alpha_1 \Delta V_{BE}}{V_{BE} + a_1 \Delta V_{BE}}, \quad (3.72)$$

$$\mu_2 = \frac{\alpha_2 \Delta V_{BE}}{V_{BE} + a_2 \Delta V_{BE}}. \quad (3.73)$$

As before, the coefficients A and B take care of scaling to degrees Celsius. The parameters α_1 , α_2 and c are chosen such that the curvature over the desired operating range is minimized.

This specific example is simple to implement if the two required AD conversions are performed by time-multiplexing a single AD converter. The analog front-end circuitry can then be the same for both conversions, only the gain of ΔV_{BE} has to be made switchable. If desired, gain factor α_1 can be chosen such that the equivalent reference voltage of the first conversion is a bandgap reference voltage (i.e. $V_{BE} + \alpha_1 \Delta V_{BE}$ has a nominally zero temperature coefficient). The first conversion then yields an uncorrected temperature reading, while the second conversion provides the curvature correction.

Figure 3.25a shows the ADC inputs for this example as a function of temperature, for various choices of α_2 . For each of these choices, there is an optimal value for the scale factor c such that the overall curvature is minimized. For all choices in this example, the optimal value for c is about 0.05, which shows that the second conversion acts as a small correction on the result of the first (and, therefore, can possibly be performed at a lower resolution). The resulting residual non-linearity is shown in Figure 3.25b. For all 5 examples, the residual non-linearity is well below 0.1°C , while for two combinations ($\alpha_2 = 3$, $c = 0.050$ and $\alpha_2 = 4$, $c = 0.050$) it is even in the order of 0.01°C . In practice, other non-idealities than systematic curvature will then be dominant.

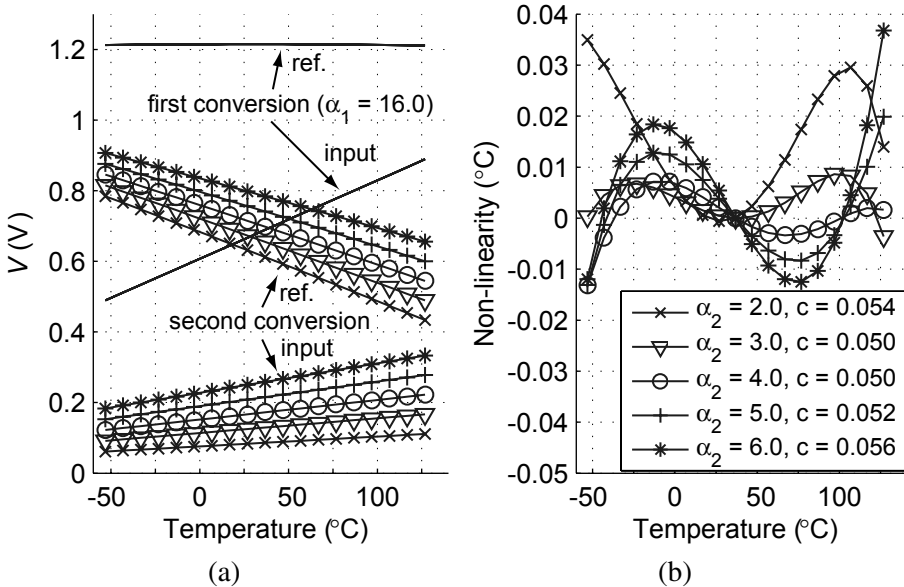


Figure 3.25. Simulation of higher-order ratiometric curvature correction: (a) ADC inputs for the first and second conversion; (b) resulting non-linearity.

3.5.6 Other Curvature-Correction Techniques

So far, curvature correction techniques have been discussed that are either based on the ratiometric transfer of an ADC, or on the correction of a bandgap voltage reference. Non-linearity can also be corrected for in the digital domain, or at the system level.

Curvature correction in the digital domain boils down to modifying the digital circuitry that converts the output μ of the ADC to a temperature reading. So far, this conversion was assumed to consist of linear scaling, as expressed by (3.7). To correct for curvature, it should introduce an inverse non-linearity that cancels the non-linearity resulting from curvature, using, for instance, a polynomial or piece-wise linear transfer function [17]. A piece-wise linear transfer can be implemented efficiently in the decimation filter of a sigma-delta ADC [53]. This will be explained in more detail in Section 4.5.4.

System-level curvature correction involves both the analog and the digital circuitry, as illustrated in Figure 3.26. A parameter in the analog front-end is adjusted by the digital circuitry. This parameter can be, for instance, the bias current of the transistor that generates V_{BE} (which may already be digitally programmable for trimming purposes), or the gain factor α . The parameter is adjusted in a temperature-dependent way so as to compensate for the curvature. Since the required adjustment typically does not change very rapidly with temperature, a coarse estimate of the temperature is sufficient to establish the

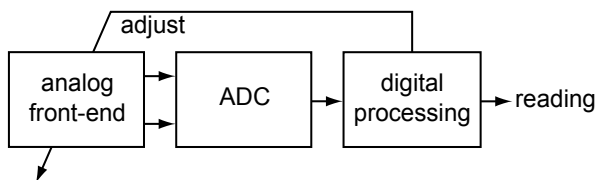


Figure 3.26. System-level curvature correction: feedback from the digital processing to the analog front-end.

adjustment. This estimate can be derived from a previous temperature reading, or from an initial reading without curvature correction.

3.6 Compensation for Finite Current-Gain

So far, the effects of the finite current-gain of practical bipolar transistors have been ignored. The base-emitter voltage V_{BE} , however, depends on the current gain, since the substrate pnp transistor used for generating V_{BE} is biased via its emitter, while V_{BE} is determined by the collector current.

3.6.1 Errors due to Finite Current-Gain

In the previous chapter, it was shown that the finite current-gain of a diode-connected transistor affects both the curvature of V_{BE} (see Figure 2.5), and increases the variation in V_{BE} due to processing spread (see Figure 2.10). The additional curvature can be compensated for using the techniques discussed in the previous section. The additional spread, however, presents a more serious problem, as it has a non-PTAT nature, and can therefore not be trimmed out using a calibration at one temperature.

The resulting errors in V_{BE} are larger for smaller nominal values of the current gain β_{F0} , and therefore become more important with every new process generation. For $\beta_{F0} = 5$, for instance, $\pm 10\%$ current-gain spread results in non-trimmable errors of ± 0.25 mV at the high end of the temperature range (see Figure 2.10), which, in turn, lead to temperature errors of almost ± 0.1 °C. For precision sensors, it is therefore important to somehow correct for these errors.

3.6.2 Current-Gain-Dependent Biasing

Obviously, errors in V_{BE} as a result of finite current-gain can be eliminated by controlling the collector current I_C of the transistor rather than its emitter current I_E . However, since the collector of substrate pnp transistors is grounded, this can only be done indirectly, e.g. by measuring the base current I_B and

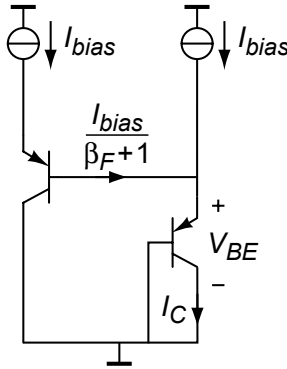


Figure 3.27. Compensation for finite current-gain by adding the base current of an auxiliary transistor to the emitter current of the primary transistor.

regulating the emitter current in a feedback loop in such a way that $I_C = I_E - I_B$ equals a desired value [54].

Alternatively, the error can be significantly reduced by adding the base current of an auxiliary transistor to the emitter current of the primary transistor, as shown in Figure 3.27 [33]. The collector current I_C of the primary transistor is then

$$\begin{aligned} I_C &= \frac{\beta_F}{\beta_F + 1} \left(1 + \frac{1}{\beta_F + 1} \right) I_{bias} \\ &= \frac{\beta_F(\beta_F + 2)}{\beta_F(\beta_F + 2) + 1} I_{bias} = \frac{\beta'_F}{\beta'_F + 1} I_{bias}. \end{aligned} \quad (3.74)$$

This feed-forward compensation results in an effective current-gain $\beta'_F = \beta_F(\beta_F + 2)$ and thus reduces the problem significantly. However, it is not suited for low-voltage operation because the base-emitter voltage of the auxiliary transistor is stacked on top of that of the primary transistor.

A simple but very effective solution to the finite current-gain problem is to use the modified PTAT/R bias circuit of Figure 3.28 [55]. Since there are no stacked base-emitter voltages in this circuit, it can operate at low supply voltages. It ensures that the collector current I_C of transistor Q is PTAT/R, rather than its emitter current. As a result, the generated V_{BE} is independent of β_F .

Compared to the PTAT/R bias circuit of Figure 3.13, a resistor R_{bias}/m has been added in series with the base of Q_{B2} . Since the feedback ensures that the input voltage of the opamp is zero, we have

$$V_{BE,B1} + I_{bias} R_{bias} = m I_{bias} \frac{1}{1 + \beta_F} \frac{R_{bias}}{m} + V_{BE,B2}. \quad (3.75)$$

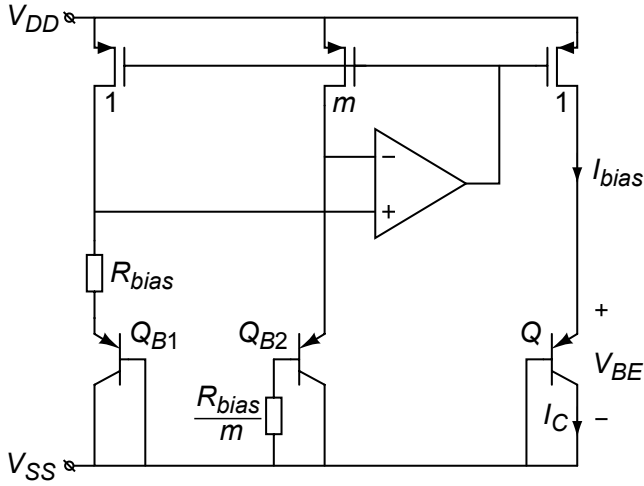


Figure 3.28. Modified PTAT/R bias circuit for compensation of the current-gain dependency of V_{BE} .

Solving for I_{bias} , we find

$$I_{bias} = \frac{1 + \beta_F}{\beta_F} \frac{V_{BE,B2} - V_{BE,B1}}{R_{bias}} = \frac{1 + \beta_F}{\beta_F} \frac{\Delta V_{BE,bias}}{R_{bias}}. \quad (3.76)$$

If this current is now applied to the emitter of transistor Q , its base-emitter voltage becomes

$$V_{BE} = \frac{kT}{q} \ln \left(\frac{I_{bias} \beta_F}{I_S \beta_F + 1} \right) = \frac{kT}{q} \ln \left(\frac{\Delta V_{BE,bias}}{R_{bias} I_S} \right), \quad (3.77)$$

which is independent of β_F , and therefore also independent of any spread of β_F .

In practice, the base resistance of Q_{B2} adds to the resistor R_{bias}/m . For perfect compensation, the total resistance in series with the (intrinsic) base of Q_{B2} has to be m times smaller than that in series with the emitter of Q_{B1} . Base-resistance results in an error in this ratio, so that V_{BE} will remain slightly dependent of β_F . This error can be partially corrected for by reducing the size of the resistor in series with the base of Q_{B2} by the nominal value of the base resistance. Residual errors are then caused by mismatch in temperature dependency and spread. Further errors may result from mismatch between the current gains of transistors Q , Q_{B1} and Q_{B2} , but these errors are expected to be negligible.

Figure 3.29 shows the simulated change in V_{BE} as a result of $\pm 10\%$ spread of β_F , for the uncompensated situation, for compensation using an auxiliary

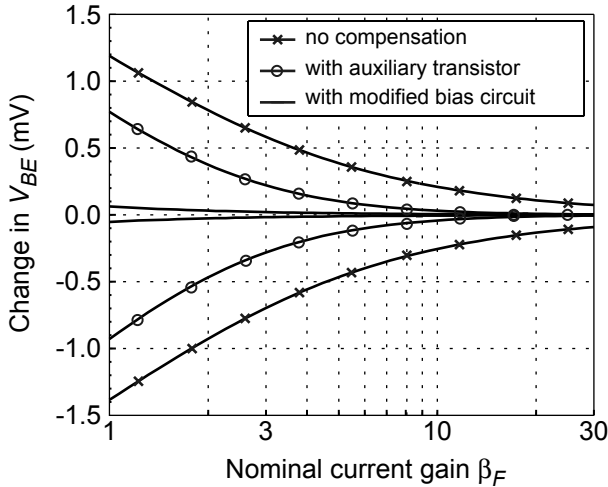


Figure 3.29. Simulated change in V_{BE} resulting from $\pm 10\%$ spread of β_F , as a function of the nominal value of β_F ($m = 8$, $R_{bias} = 200 \text{ k}\Omega$, $R_B = 500 \Omega$, $T = 300 \text{ K}$).

transistor (Figure 3.27), and for the modified PTAT/R bias circuit (Figure 3.28). For current gains larger than 10, the performance of the compensation using an auxiliary transistor and the modified PTAT/R bias circuit is comparable. As expected, the modified PTAT/R bias circuit performs much better for lower current-gains: the spread is reduced by more than 10 times to less than 0.1 mV . This means that the errors after PTAT trimming (in Figure 2.10) are also reduced by a factor 10 and become negligible. The residual dependency at low current-gains is caused by the base resistance of Q_{B2} .

3.7 Series-Resistance Compensation

3.7.1 Errors due to Series Resistances

In Section 2.7, it has been shown that the combined effect of series resistances associated with the base, the emitter and the collector of a diode-connected transistor can be modelled as a resistor R_S in series with the emitter. With the relatively large base resistance and low current gain of substrate pnp transistors, values of typically a few tens of ohm are found.

The voltage drop across series resistances modifies V_{BE} and ΔV_{BE} as follows:

$$V_{BE} = V_{BE}|_{R_S=0} + I_2 R_S, \quad (3.78)$$

$$\Delta V_{BE} = \Delta V_{BE}|_{R_S=0} + I_1 \left(p - \frac{1}{r} \right) R_S, \quad (3.79)$$

where I_1 and I_2 are the currents used for generating ΔV_{BE} and V_{BE} , respectively, and p and r are, as before, the bias-current ratio and the emitter-area ratio (see Figure 3.1).

The associated temperature errors can be found by multiplying the voltage drops by the sensitivity of D_{out} to errors in V_{BE} and ΔV_{BE} . Using equation (3.9), the temperature error due to series resistance of the transistor that generates V_{BE} is found to be

$$\Delta T = -\frac{I_2 R_S}{V_{REF}} T. \quad (3.80)$$

Since $R_S I_2$ is typically in the order of a few tens of μV , while V_{REF} is approximately 1.2 V, this error is usually negligible.

The temperature error due to series resistances of the transistors that generate ΔV_{BE} can be calculated using equation (3.10):

$$\Delta T = \frac{R_S I_1}{V_{REF}} \alpha \left(p - \frac{1}{r} \right) (A - T). \quad (3.81)$$

Assuming that the current levels used for generating ΔV_{BE} are in the same order of magnitude as those used for generating V_{BE} , this error is about α times larger, and therefore not necessarily negligible. Since the voltage drop across series resistances is unlikely to be PTAT, it will not be corrected for by trimming, and should be reduced to negligible levels by design. This can be done either by optimizing the transistor geometry, by reducing the bias currents, or by using one of the techniques discussed in the following sections.

3.7.2 Instantaneous Compensation

The voltage drop across series resistances of a transistor can be compensated for by generating an equal voltage drop across an auxiliary device that mimics the series resistance. This can be a resistor which is designed to be nominally equal to the series resistance [56], but matching over temperature and process spread will then generally be poor.

A better solution is to use identical transistors and exploit the fact that the intrinsic base-emitter voltages scales logarithmically with current, while the voltage drop across series resistances scales linearly. This technique was originally developed for use in translinear circuits, such as multipliers and dividers [57]. Figure 3.30a shows how three identical substrate pnp transistors can be used to generate a base-emitter voltage V_{BE} free from series-resistance errors.

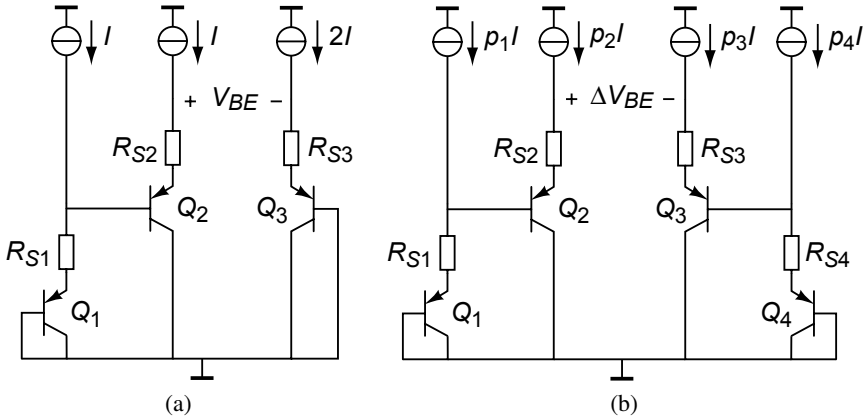


Figure 3.30. Instantaneous series-resistance compensation for (a) V_{BE} and (b) ΔV_{BE} .

Two transistors Q_1 and Q_2 are connected in series and are both biased at a current I , while the third transistor Q_3 is biased at $2I$. The differential voltage V_{BE} is then:

$$\begin{aligned}
 V_{BE} &= V_{BE1} + V_{BE2} - V_{BE3} \\
 &= \frac{kT}{q} \ln \left(\frac{I}{2I_S} \right) + I(R_{S1} + R_{S2} - 2R_{S3}) \\
 &= \frac{kT}{q} \ln \left(\frac{I}{2I_S} \right), \tag{3.82}
 \end{aligned}$$

which equals the base-emitter voltage of a transistor without series resistances biased at a current $I/2$, assuming the base currents and the Early effect can be ignored, and the series resistances of the three transistors are equal.

In a similar way, a ΔV_{BE} can be generated which is free from series-resistance errors (Figure 3.30b). Under the same assumptions,

$$\Delta V_{BE} = V_{BE1} + V_{BE2} - V_{BE3} - V_{BE4} \tag{3.83}$$

$$= \frac{kT}{q} \ln \left(\frac{p_1 p_2}{p_3 p_4} \right) + (p_1 + p_2 - p_3 - p_4) \cdot I \cdot R_S, \tag{3.84}$$

which shows that the series-resistance terms cancel provided that $p_1 + p_2 - p_3 - p_4 = 0$. An example of a combination that satisfies this requirement is $p_1 = 3$, $p_2 = 4$, $p_3 = 1$, and $p_4 = 6$.

A disadvantage of the techniques of Figure 3.30 is that they require a supply voltage of at least two base-emitter voltages and a saturation voltage. Care should be taken that the transistors are operated at current levels where

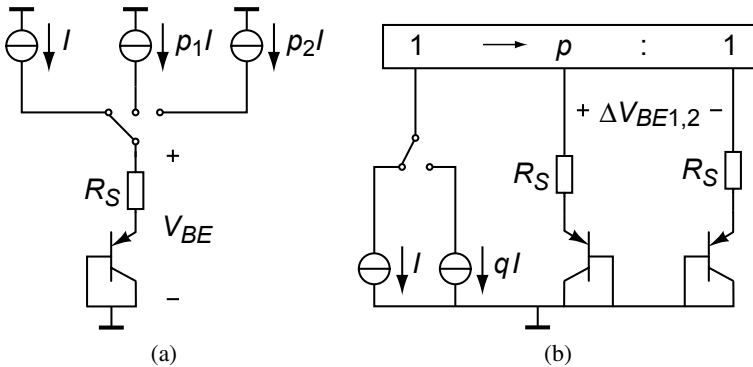


Figure 3.31. Sequential series-resistance compensation using (a) a single transistor and (b) two transistors.

their resistances are current-independent. Current crowding, which occurs at high currents and makes the base resistance current-dependent [58], should therefore be avoided.

3.7.3 Sequential Compensation

If a low supply voltage is required, or if the errors due to base currents and the Early effect cannot be ignored, a sequential approach can be used, in which a number of bias currents are successively applied to one or two transistors, and the resulting voltages are sampled and combined. An implementation with a single transistor is shown in Figure 3.31a [59]. If two bias currents I and $p_1 I$ are successively applied to a diode-connected pnp transistor, the resulting base-emitter voltages can be combined as follows to eliminate series-resistance terms:

$$\frac{p_1}{p_1 - 1} V_{BE}(I) - \frac{1}{p_1 - 1} V_{BE}(p_1 I) = \frac{kT}{q} \ln \left(\frac{I}{I_S} \right) - \frac{1}{p_1 - 1} \frac{kT}{q} \ln(p_1). \quad (3.85)$$

Taking for instance $p_1 = 2$, this becomes

$$2 \cdot V_{BE}(I) - V_{BE}(2I) = \frac{kT}{q} \ln \left(\frac{I}{2I_S} \right). \quad (3.86)$$

To generate a ΔV_{BE} that is free of series-resistance terms, three bias currents I , $p_1 I$ and $p_2 I$ are needed. Using, for example, $p_1 = 3$ and $p_2 = 9$, the following combination can be used:

$$4 \cdot V_{BE}(3I) - 3 \cdot V_{BE}(I) - V_{BE}(9I) = \frac{kT}{q} \ln(9). \quad (3.87)$$

An implementation with two transistors is shown in Figure 3.31b. A current mirror biases these transistors at a $p : 1$ current ratio, while the unit current can

be switched between I and qI . This results in two differences:

$$\Delta V_{BE1} = V_{BE}(pI) - V_{BE}(I) = \frac{kT}{q} \ln(p) + (p - 1) \cdot I \cdot R_S, \quad (3.88)$$

$$\Delta V_{BE2} = V_{BE}(pqI) - V_{BE}(qI) = \frac{kT}{q} \ln(p) + (p - 1) \cdot q \cdot I \cdot R_S. \quad (3.89)$$

Compensation is now obtained from the combination

$$\Delta V_{BE} = q\Delta V_{BE1} - \Delta V_{BE2} = \frac{kT}{q} (q - 1) \ln(p). \quad (3.90)$$

If, for instance, $q = 2$, this comes down to doubling the bias currents when generating ΔV_{BE2} and then taking $2\Delta V_{BE1} - \Delta V_{BE2}$.

The two-transistor implementation has the advantage that the dynamic range of the readout circuit can be smaller than in a single transistor implementation, where a full V_{BE} has to be processed. Moreover, the matching of the $p : 1$ mirror ratio can be controlled independently of the $q : 1$ current-source ratio. An error Δp in the former ratio affects ΔV_{BE} in the same way as discussed in Section 3.2, and can be compensated for using dynamic element matching (DEM). An error Δq in the latter ratio, however, only affects the series-resistance terms in (3.88) and (3.89). It results in a imperfect compensation, which is equivalent to a residual series resistance of $\Delta q/q \cdot R_S$. With an error $\Delta q/q$ of 1%, the voltage drop across series resistances is therefore still reduced by a factor 100, which is usually more than enough. Therefore, no DEM is required for the $q : 1$ ratio, which simplifies the design. Note that the factor $(q - 1)$ in (3.90), in contrast, needs to be as accurate as the $p : 1$ ratio; this factor is however implemented in the readout circuitry; its accuracy is unrelated to that of the $q : 1$ bias current ratio.

Note that all described techniques inevitably decrease the signal-to-noise ratio compared to a situation without series-resistance compensation: the noise of at least one extra transistor is added, while no signal gain comes in return. It may be necessary to increase the bias currents to compensate for this increased noise.

3.8 Conclusions

This chapter has described how substrate pnp transistors can be used for accurate ratiometric temperature measurement. The difference in base-emitter voltage ΔV_{BE} between two substrate pnp transistors operated at different current densities is proportional to absolute temperature (PTAT), while the base-emitter voltage V_{BE} of a single transistor is complementary to absolute temperature (CTAT). It has been shown how ΔV_{BE} and V_{BE} can be combined to generate two voltages of which the ratio is an accurate function of temperature.

An analog-to-digital converter (ADC) is used to compute this ratio, and, with appropriate digital scaling, produce a temperature reading in degrees Celsius.

The sensitivity of ΔV_{BE} , in principle, only depends on the current-density ratio, which in turn depends on device matching. With 0.1% mismatch, as can be expected from precise layout, device mismatches lead to temperature errors larger than 0.1°C . Dynamic element matching (DEM) can be used to average out these mismatches, so that with an initial matching of 1%, temperature errors well below 0.1°C can be obtained.

The base-emitter voltage V_{BE} , in contrast, is process-dependent. It exhibits a PTAT spread as a result of variations in the transistor's saturation current I_S . Using a calibration at one temperature, this spread can be trimmed out, but only if other variations, such as variations in the current used to bias the transistor, do not disturb the PTAT character of the spread.

Any bias current generated on chip is derived from an on-chip bias voltage using a bias resistor. A PTAT voltage is a good choice for the bias voltage, given its high initial accuracy. The spread of the resulting PTAT/R bias current is then determined by the bias resistor. Spread of the nominal value of this resistor is not a problem, since it just adds to the PTAT spread of V_{BE} , but spread in the resistor's temperature behavior should be minimized, as it leads to spread in the curvature of V_{BE} . Therefore, the use of a bias resistor with a small or at least a reproducible temperature coefficient is advisable. Requirements for the circuit parameters of the bias circuit, such as offset, matching and loop gain, have been derived.

Trimming of V_{BE} boils down to adding a programmable PTAT voltage to it. The required magnitude of this voltage can be found by calibration, i.e. by comparing a reading of the sensor with that of an accurate reference thermometer. The actual adjustment of V_{BE} can be implemented in the voltage domain, by implementing a programmable PTAT voltage source, or in the current domain, by making the transistor's bias current or its emitter area programmable. To obtain a high trimming resolution without using a large number of switchable unit elements, modulation techniques can be applied. Alternatively, the trimming can be implemented in the digital domain. To obtain the same effect as a PTAT adjustment of V_{BE} , a non-linear adjustment in the digital processing would be needed. Fortunately, a correlated offset and gain adjustment can mimic the effect of a PTAT adjustment to within $\pm 0.1^\circ\text{C}$. Irrespective of the chosen trimming technique, non-volatile memory is needed to store the trim setting. In CMOS, floating-gate technology such as EPROM is usually applied for this purpose. Trimming after packaging can, to some extent, correct for stress-induced changes. As a result of the non-PTAT temperature dependency of these changes, however, the correction will only be effective close to the calibration temperature.

The curvature of V_{BE} results in a systematic non-linearity of the sensor. If V_{BE} and ΔV_{BE} are used to make a conventional bandgap voltage reference for the ADC (i.e. with a nominally zero temperature coefficient), this non-linearity amounts to about 1.0°C (depending on the curvature parameters of the transistor). The various curvature-correction techniques for bandgap references that can be found in literature can be applied to reduce this non-linearity. A more efficient way of reducing the non-linearity is to exploit the ratiometric nature of the temperature measurement: by using a reference with a small positive temperature coefficient, a compensating second-order non-linearity can be introduced that reduces the overall non-linearity to less than 0.2°C . This ‘ratiometric’ curvature correction can be extended to higher-orders by combining the results of multiple AD conversions which use different temperature-dependent references. In this way, an overall non-linearity well below 0.1°C can be obtained. Alternatively, curvature can also be corrected for in digital post-processing, or at the system-level.

Substrate pnp transistors are biased via their emitter, because their collector (the substrate) is inaccessible. As a result, their base-emitter voltage is not only a function of the transistor’s saturation current and the bias current, but also of the transistor’s current gain. Since the current gain of substrate pnp transistors is relatively small (and gets smaller with every CMOS generation), spread of the current gain results in a significant spread of V_{BE} . Because the current gain is temperature dependent, this spread does not have a PTAT characteristic and therefore cannot be trimmed out completely. Since the associated errors can be larger than 0.1°C , it is important to compensate for finite current-gain. With a simple modification, a PTAT/R bias circuit can be used to generate a current-gain-dependent bias current, which, when applied to the emitter of a diode-connected substrate pnp transistor, results in a PTAT/R collector current. The resulting base-emitter voltage is then independent of the current gain.

The series resistance associated with substrate pnp transistors represents a final source of inaccuracy. The effective series resistance seen at the emitter of a substrate pnp transistor is dominated by its base resistance, but also includes its emitter resistance and any interconnect resistances. The voltage drop across this series resistance adds to V_{BE} and ΔV_{BE} . Spread of the series resistance, due to its temperature dependency, cannot be trimmed out along with the PTAT spread of V_{BE} . Therefore, the series resistance has to be minimized, by optimization of the transistor’s geometry, or by reducing the current levels. If this is not sufficient, various compensation techniques can be applied that make use of the fact that the voltage drop across series resistances scales linearly with current, while the intrinsic base-emitter voltage scales logarithmically. Thus, by operating transistors at different current densities and combining their base-emitter voltages, series resistance terms can be eliminated.

The voltages V_{BE} and ΔV_{BE} generated using the techniques outlined in this chapter can be used to realize temperature sensors with inaccuracies of ± 0.1 °C. The design of the required ratiometric ADC, which turns these voltages into a digital temperature reading, is the topic of the next chapter.

References

- [1] G. C. M. Meijer, "Integrated circuits and components for bandgap references and temperature transducers," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, Mar. 1982.
- [2] A. Bakker and J. H. Huijsing, *High-Accuracy CMOS Smart Temperature Sensors*. Boston: Kluwer Academic Publishers, 2000.
- [3] J. V. Nicholas and D. R. White, *Traceable Temperatures*. Chichester, England: John Wiley & Sons, 1994.
- [4] A. Hastings, *The art of analog layout*. New Jersey: Prentice Hall, 2001.
- [5] K. B. Klaassen, "Digitally controlled absolute voltage division," *IEEE Transactions on Instrumentation and Measurement*, vol. 24, no. 2, pp. 106–112, June 1975.
- [6] R. J. van der Plassche, "Dynamic element matching for high-accuracy monolithic D/A converters," *IEEE Journal of Solid-State Circuits*, vol. SC-11, no. 6, pp. 795–800, Dec. 1976.
- [7] G. C. M. Meijer, G. Wang, and F. Fruett, "Temperature sensors and voltage references implemented in CMOS technology," *IEEE Sensors Journal*, vol. 1, no. 3, pp. 225–234, Oct. 2001.
- [8] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*. Chichester, England: John Wiley & Sons, 2001.
- [9] B. Razavi, *Design of Analog CMOS Integrated Circuits*. New York: McGraw-Hill, 2001.
- [10] R. W. Dutton and D. A. Divekar, "Bipolar models for statistical IC design," in *Process and device modeling for integrated circuit design*, F. van de Wiele *et al.*, Eds. Addison-Wesley, 1977, pp. 461–517.
- [11] B. Gilbert, "Monolithic voltage and current references: theme and variations," in *Analog Circuit Design*, J. H. Huijsing *et al.*, Eds. Boston: Kluwer Academic Publishers, 1996, pp. 269–352.
- [12] J. Michejda and S. K. Kim, "A precision CMOS bandgap reference," *IEEE Journal of Solid-State Circuits*, vol. SC-19, no. 6, pp. 1014–1021, Dec. 1984.
- [13] R. J. Fronen, "Band-gap reference current source with compensation for saturation current spread of bipolar transistors," U.S. Patent 5 581 174, Dec. 2, 1994.
- [14] R. Amador, A. Polanco, H. Hernández, E. González, and A. Nagy, "Technological compensation circuit for accurate temperature sensor," *Sensors and Actuators*, vol. 69, no. 2, pp. 172–177, Aug. 1998.

- [15] R. Amador, A. Polanco, H. Hernández, E. González, and A. Nagy, "Reducing V_{BE} wafer spread of bipolar transistor via a compensation circuit," *Electronics Letters*, vol. 28, no. 15, pp. 1378–1379, July 1992.
- [16] M. A. P. Pertijs and J. H. Huijsing, "Bitstream trimming of a smart temperature sensor," in *Proc. IEEE Sensors*, Oct. 2004, pp. 904–907.
- [17] G. v. d. Horn and J. H. Huijsing, *Integrated Smart Sensors: Design and Calibration*. Boston: Kluwer Academic Publishers, 1998.
- [18] F. Fruett and G. C. M. Meijer, *The Piezjunction Effect in Silicon Integrated Circuits and Sensors*. Boston: Kluwer Academic Publishers, May 2002.
- [19] B. Abesingha, G. A. Rincón-Mora, and D. Briggs, "Voltage shift in plastic-packaged bandgap references," *IEEE Transactions on Circuits and Systems—Part II: Analog and Digital Signal Processing*, vol. 49, no. 10, pp. 681–685, Oct. 2002.
- [20] A. Elshabini-Riad and I. A. Bhutta, "Lightly trimming the hybrids," *IEEE Circuits and Devices Magazine*, vol. 9, no. 4, pp. 30–34, July 1993.
- [21] J. A. Babcock, D. W. Feldbaumer, and V. M. Mercier, "Polysilicon resistor trimming for packaged integrated circuits," in *Proc. IEDM*, Dec. 1993, pp. 247–250.
- [22] G. Erdi, "A precision trim technique for monolithic analog circuits," *IEEE Journal of Solid-State Circuits*, vol. SC-10, no. 6, pp. 412–416, Dec. 1975.
- [23] J. Teichmann, K. Burger, W. Hasche, J. Herrfurth, and G. Täschner, "One time programming (OTP) with zener diodes in CMOS processes," in *Proc. ESSDERC*, Sept. 2003, pp. 433–436.
- [24] G. A. Rincón-Mora, *Voltage References*. Piscataway, New York: IEEE Press, 2002.
- [25] M. de Wit, K.-S. Tan, and R. K. Hester, "A low-power 12-b analog-to-digital converter with on-chip precision trimming," *IEEE Journal of Solid-State Circuits*, vol. 28, no. 4, pp. 455–461, Apr. 1993.
- [26] A. F. Murray and L. W. Buchan, "A user's guide to non-volatile, on-chip analogue memory," *IEE Electronics & Communication Engineering Journal*, vol. 10, no. 2, pp. 53–63, Apr. 1998.
- [27] E. Säckinger and W. Guggenbühl, "An analog trimming circuit based on a floating-gate device," *IEEE Journal of Solid-State Circuits*, vol. 23, no. 6, pp. 1437–1440, Dec. 1988.
- [28] G. Wang and G. C. M. Meijer, "Temperature characteristics of bipolar transistors fabricated in CMOS technology," *Sensors and Actuators*, vol. 87, pp. 81–89, Dec. 2000.
- [29] G. A. Rincón-Mora and P. E. Allen, "A 1.1-V current-mode and piecewise-linear curvature-corrected bandgap reference," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 10, pp. 1551–1554, Oct. 1998.
- [30] "MAX6325 data sheet," Maxim Int. Prod., Dec. 2003, www.maxim-ic.com.
- [31] "AD588 data sheet," Analog Devices Inc., Feb. 2003, www.analog.com.

- [32] A. P. Brokaw, "A simple three-terminal IC bandgap reference," *IEEE Journal of Solid-State Circuits*, vol. SC-9, no. 6, pp. 388–393, Dec. 1974.
- [33] B.-S. Song and P. R. Gray, "A precision curvature-compensated CMOS bandgap reference," *IEEE Journal of Solid-State Circuits*, vol. SC-18, no. 6, pp. 634–643, Dec. 1983.
- [34] I. M. Filanovsky and Y. F. Chan, "BiCMOS cascaded bandgap voltage reference," in *Proc. Midwest Symposium on Circuits and Systems*, vol. 2, Aug. 1996, pp. 943–946.
- [35] J. T. Sundby, "Low voltage CMOS bandgap with new trimming and curvature correction methods," U.S. Patent 5 325 045, June 28, 1994.
- [36] C. Falconi, A. D'Amico, C. D. Natale, and M. Faccio, "Low cost curvature correction of bandgap references for integrated sensors," *Sensors and Actuators*, vol. 117, no. 1, pp. 127–136, Jan. 2005.
- [37] S. R. Lewis and A. P. Brokaw, "Curvature correction of bipolar bandgap references," U.S. Patent 4 808 908, Feb. 28, 1989.
- [38] J. M. Audy, "Bandgap voltage reference circuit and method with low TCR resistor in parallel with high TCR and in series with low TCR portions of tail resistor," U.S. Patent 5 291 122, Mar. 1, 1994.
- [39] J. M. Audy, "3rd order curvature corrected bandgap cell," in *Proc. Midwest Symposium on Circuits and Systems*, vol. 1, Aug. 1995, pp. 397–400.
- [40] K. N. Leung, P. K. T. Mok, and C. Y. Leung, "A 2-V 23- μ A 5.3-ppm/ $^{\circ}$ C curvature-compensated CMOS bandgap voltage reference," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 3, pp. 561–564, Mar. 2003.
- [41] R. J. Widlar, "Low voltage techniques," *IEEE Journal of Solid-State Circuits*, vol. SC-13, no. 6, pp. 838–846, Dec. 1978.
- [42] G. C. M. Meijer, P. C. Schmale, and K. van Zalinge, "A new curvature-corrected bandgap reference," *IEEE Journal of Solid-State Circuits*, vol. SC-17, no. 6, pp. 1139–1143, Dec. 1982.
- [43] S. L. Lin and C. A. T. Salama, "A $V_{BE}(T)$ model with application to bandgap reference design," *IEEE Journal of Solid-State Circuits*, vol. SC-20, no. 6, pp. 1283–1285, Dec. 1985.
- [44] P. Malcovati, F. Maloberti, C. Fiocchi, and M. Pruzzi, "Curvature-compensated BiCMOS bandgap with 1-V supply voltage," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 7, pp. 1076–1081, July 2001.
- [45] C. Hagleitner, "CMOS single-chip gas detection system comprising capacitive, calorimetric and mass-sensitive microsensors," Ph.D. dissertation, Swiss Federal Institute of Technology, Zurich, Switzerland, 2002.
- [46] C. R. Palmer and R. C. Dobkin, "A curvature corrected micropower voltage reference," in *Dig. Techn. Papers ISSCC*, Feb. 1981, pp. 58–59.
- [47] I. Lee, G. Kim, and W. Kim, "Exponential curvature-compensated BiCMOS bandgap references," *IEEE Journal of Solid-State Circuits*, vol. 29, no. 11, pp. 1396–1403, Nov. 1994.

- [48] S.-Y. Chin and C.-Y. Wu, "A new type of curvature-compensated CMOS bandgap voltage references," in *Proc. Int. Symp. on VLSI Techn.*, May 1991, pp. 398–402.
- [49] O. Salminen and K. Halonen, "The higher order temperature compensation of bandgap voltage references," in *Proc. ISCAS*, vol. 3, May 1992, pp. 1388–1391.
- [50] G. C. M. Meijer *et al.*, "A three-terminal integrated temperature transducer with micro-computer interfacing," *Sensors and Actuators*, vol. 18, pp. 195–206, June 1989.
- [51] M. A. P. Pertijs, A. Bakker, and J. H. Huijsing, "A high-accuracy temperature sensor with second-order curvature correction and digital bus interface," in *Proc. ISCAS*, May 2001, pp. 368–371.
- [52] M. A. P. Pertijs, A. Bakker, and J. H. Huijsing, "Non-linear signal correction," U.S. Patent 6 456 145, Sept. 24, 2002.
- [53] P. Malcovati, C. A. Leme, P. O'Leary, F. Maloberti, and H. Baltes, "Smart sensor interface with A/D conversion and programmable calibration," *IEEE Journal of Solid-State Circuits*, vol. 29, no. 8, pp. 963–966, Aug. 1994.
- [54] I. M. Filanovsky and S. S. Cai, "BiCMOS bandgap voltage reference with base current compensation," *Int. J. Electronics*, vol. 81, no. 5, pp. 565–570, 1996.
- [55] M. A. P. Pertijs and J. H. Huijsing, "Bias circuits," U.K. Patent Application 0 420 484.8, 2005.
- [56] B. Song and P. R. Gray, "A precision curvature-compensated CMOS bandgap reference," in *Dig. Techn. Papers ISSCC*, Feb. 1983, pp. 240–241.
- [57] I. Opris, "Series resistance compensation in translinear circuits," *IEEE Transactions on Circuits and Systems—Part I: Fundamental Theory and Applications*, vol. 45, no. 1, pp. 91–94, Jan. 1998.
- [58] P. A. H. Hart, Ed., *Bipolar and Bipolar-MOS Integration*. Amsterdam, The Netherlands: Elsevier, 1994.
- [59] J. M. Audy and B. Gilbert, "Multiple sequential excitation temperature sensing method and apparatus," U.S. Patent 5 195 827, Mar. 4, 1993.

Chapter 4

SIGMA-DELTA ANALOG-TO-DIGITAL CONVERSION

This chapter discusses the design of the analog-to-digital converter (ADC) of a precision smart temperature sensor. This ADC converts the voltages V_{BE} and ΔV_{BE} (generated using the techniques introduced in the previous chapter) to a digital temperature reading. The chapter starts with an overview of the requirements that have to be met in this application. After a brief overview of different types of ADCs, sigma-delta ($\Sigma\Delta$) ADCs are shown to be particularly suited for the narrow bandwidth signals found in temperature sensors. The system-level design of first- and second-order $\Sigma\Delta$ modulators and the associated decimation filters is discussed. Since dynamic error correction techniques (such as dynamic element matching) are needed to accurately generate V_{BE} and ΔV_{BE} , special attention is paid to the filtering of the associated dynamic error signals.

4.1 Introduction

An ADC is a key building block in a smart temperature sensor. In the previous chapter, it was described how temperature can be determined from a ratio of voltages derived from bipolar transistors. The ADC in a temperature sensor serves to digitize this ratio. After a brief overview of ADC types, it will be shown that the class of charge-balancing ADCs, and specifically $\Sigma\Delta$ ADCs, are well suited for this purpose. But first, the requirements on such ADCs will be reviewed.

4.1.1 Requirements

The ADC in a temperature sensor has to produce a digital output D_{out} from the voltages generated by the front-end circuitry. As described in Section 3.1.1, these voltages are typically a base-emitter voltage V_{BE} and a difference in base-

emitter voltage ΔV_{BE} (see Figure 3.1). They can be combined, as expressed by (3.6), to obtain a ratio μ that is an accurate function of temperature:

$$\mu = \frac{\alpha \cdot \Delta V_{BE}}{V_{BE} + \alpha \cdot \Delta V_{BE}}. \quad (4.1)$$

In this ratio, the numerator is proportional to absolute temperature (PTAT), while the denominator is temperature-independent (ignoring curvature effects for simplicity), so that the ratio μ is PTAT. As expressed by (3.7), some scaling is required to obtain output data D_{out} in degrees Celsius:

$$D_{out} = A \cdot \mu + B. \quad (4.2)$$

Accuracy

In Section 3.1.3, accuracy requirements (3.12) and (3.13) for the ADC's input signals V_{BE} and ΔV_{BE} have been derived:

$$|V_{BE} - V_{BE,ideal}| < (3 \text{ mV} / ^\circ\text{C}) \cdot \Delta T, \quad (4.3)$$

$$|\Delta V_{BE} - \Delta V_{BE,ideal}| < \frac{3 \text{ mV} / ^\circ\text{C}}{\alpha} \cdot \Delta T. \quad (4.4)$$

These equations express the maximum error in V_{BE} and ΔV_{BE} for a given maximum temperature error ΔT over the military temperature range.

To ensure that the ADC does not add significant errors, its input-referred errors, including offset errors, gain errors and non-linearity, have to meet the same requirements. For a maximum temperature error of $\pm 0.01^\circ\text{C}$ and a value of $\alpha = 10$, for example, the maximum input-referred offset that the ADC can add to ΔV_{BE} is $\pm 3 \mu\text{V}$.

The accuracy requirements above were derived by calculating the sensitivity of D_{out} to errors in V_{BE} and ΔV_{BE} . Similarly, the required absolute accuracy of the scale factors α , A and B can be derived by determining the sensitivity of D_{out} to errors in these factors. This leads to the following requirements:

$$\left| \frac{\alpha - \alpha_{ideal}}{\alpha_{ideal}} \right| < \left(\frac{2}{3} \% / ^\circ\text{C} \right) \cdot \Delta T, \quad (4.5)$$

$$\left| \frac{A - A_{ideal}}{A_{ideal}} \right| < \left(\frac{1}{4} \% / ^\circ\text{C} \right) \cdot \Delta T, \quad (4.6)$$

$$|B - B_{ideal}| < \Delta T. \quad (4.7)$$

For example, a maximum temperature error ΔT of $\pm 0.01^\circ\text{C}$ corresponds to a maximum error of $\pm 0.0066\%$ in α , a maximum error of $\pm 0.0025\%$ in A , and a maximum error of $\pm 0.1^\circ\text{C}$ in B .

Effective Number of Bits

The resolution required at the output of the ADC depends on the requirements of the application. In general, it should be high enough to make quantization errors insignificant compared to the accuracy requirement of the sensor. For a sensor with a desired inaccuracy of $\pm 0.1^\circ\text{C}$, for instance, a resolution of $\pm 0.01^\circ\text{C}$ makes quantization errors negligible. For applications where small changes in temperature have to be measured (rather than absolute temperature), a higher resolution might be required.

In this work, the term ‘effective number of bits’ (ENOB) will be used to express the ADC’s total quantization error as a fraction of its full scale. According to the IEEE definition, an ADC has a given ENOB if its signal-to-noise and distortion ratio equals that of an ideal ADC with the same number of bits [1]. The ENOB according to this definition is a figure of merit for the AC performance of an ADC. Here, the term will be used similarly to specify the DC performance: it will be used to express the fact that an ADC’s peak quantization error over its DC input range (including offset errors, gain errors, non-linearity) equals that of an ideal ADC with the same number of bits¹. Since the quantization error of an ideal ADC is less than ± 0.5 LSB, an ENOB of n bits implies that the quantization error $D_{out} - D_{out,ideal}$ of the ADC satisfies

$$|D_{out} - D_{out,ideal}| < \frac{D_{FS}}{2^{\text{ENOB}+1}}, \quad (4.8)$$

where D_{FS} is the full-scale value. Conversely, the ENOB can be expressed as

$$\text{ENOB} = \log_2 \left(\frac{D_{FS}}{\max(|D_{out} - D_{out,ideal}|)} \right) - 1. \quad (4.9)$$

The ENOB required in a precision temperature sensor can now be calculated as follows. As discussed in Section 3.1.1, for the ratio (4.1), the extremes of the military operating temperature range of -55°C to 125°C correspond to roughly $\mu = \frac{1}{3}$ and $\mu = \frac{2}{3}$. Therefore, only about $\frac{1}{3}$ of the ADC’s full scale will be used. The full scale then corresponds to a temperature range of about 600°C . Combined with a maximum quantization error of $\pm 0.01^\circ\text{C}$, this leads to a required ENOB of 14.9 bits.

Note that requirement (4.8) only has to be met over the part of the input range that is actually used. Near the extremes of the input range, larger errors can be tolerated. As will be shown in Section 4.4, this is convenient for single-loop second-order $\Sigma\Delta$ ADCs, as they tend to produce increased quantization errors at the extremes of their input range.

¹Note that this is not equal to the integral non-linearity (INL): INL does not include offset and gain errors [1].

With a more complex combination of ΔV_{BE} and V_{BE} that uses the ADC's input range more efficiently, such as that given by equation (3.8), the full scale corresponds to about 200 °C. The ENOB required for a peak quantization error of ± 0.01 °C is then reduced to 13.3 bits, but this number has to be met over 90% of the input range (see Figure 3.3).

Bandwidth

The bandwidth of general-purpose smart temperature sensors is typically limited by the thermal properties of its package, and is in the order of 10 Hz [2]. The ADC should therefore have a similar bandwidth, and should be able to take about 10 temperature readings per second. For sensors used in a control loop, a larger bandwidth may be required for stability reasons [3]. That application will not be considered further in this work.

4.1.2 Direct versus Indirect Conversion

When an ADC digitizes its input signal, it essentially divides its input range into a number of segments, and then determines in which segment the current value of the input signal lies. Converters that literally make such a segmentation are sometimes referred to as *direct* ADCs, and include flash ADCs, successive-approximation ADCs, algorithmic ADCs, and pipeline ADCs [4]. In direct ADCs, the segmentation of the input range is based on the matching on-chip components (resistors or capacitors). As a result, the accuracy of these converters is limited to about 12 bits. This is most clear for flash ADCs, which literally divide the reference voltage using a resistive ladder, and use a set of comparators to determine in which segment the input voltage lies. The matching of the resistors then directly determines the accuracy. Direct ADCs are most suitable for high-speed low-resolution conversion. Smart temperature sensors, in contrast, typically require high resolution at a very low speed.

Converters that use an intermediate step to arrive at an output value are called *indirect* ADCs, and include dual-slope and $\Sigma\Delta$ ADCs, and ADCs based on duty-cycle, frequency or period modulation [4]. They first convert the ratio of the input voltage and the reference voltage into a time-domain ratio. In the case of a dual-slope ADC, for example, this is the ratio of two time intervals. In the case of a $\Sigma\Delta$ ADC, it is the fraction of ones in a bitstream. This time-domain ratio is converted into a final conversion result by digital circuitry that typically runs at a much higher frequency than the conversion rate, e.g. a counter or a decimation filter. Indirect ADCs do not rely on the matching of on-chip components. Instead, they trade speed for resolution, and are therefore a good match to the low speed and high resolution required in a smart temperature sensor.

For this reason, smart temperature sensors found in literature almost invariably use indirect ADCs [5–8]. The design by Tuthill is a notable exception [9].

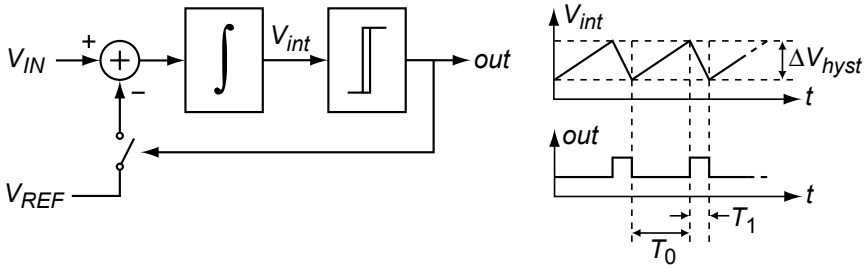


Figure 4.1. Block diagram and timing of a duty-cycle modulator.

In this design, however, a successive-approximation ADC was chosen because other input signals in addition to temperature had to be digitized that required a high conversion speed.

4.1.3 Charge Balancing

The conversion from the voltage-domain to the time-domain in an indirect ADC is usually performed by a modulator. This typically consists of at least one integrator and a comparator. Figure 4.1 shows, as an example, the block diagram and timing of a duty-cycle modulator. The input signal V_{IN} is continuously integrated, while the reference voltage is subtracted from V_{IN} if the integrator's output exceeds the comparator's threshold. As a result of the hysteresis ΔV_{hyst} of the comparator, the integrator's output V_{int} will oscillate between the thresholds of the comparator. The increase of the integrator's output during the time T_0 that the comparator's output is low, is therefore equal to its decrease during the time T_1 that the comparator's output is high. Assuming V_{IN} and V_{REF} can be considered constant, this can be expressed as

$$T_0 \cdot V_{IN} = -T_1 \cdot (V_{IN} - V_{REF}), \quad (4.10)$$

which can be rewritten as

$$\frac{T_1}{T_0 + T_1} = \frac{V_{IN}}{V_{REF}}. \quad (4.11)$$

This shows that the duty cycle of the comparator's output is equal to the ratio of V_{IN} and V_{REF} . This type of operation is sometimes referred to as *charge balancing*: the charge accumulated due to the integration of V_{IN} is balanced by the charge accumulated due to the integration of V_{REF} . Other indirect ADCs make use of this same principle.

The transfer function (4.1) of a smart temperature sensor can be implemented very elegantly using charge balancing [5, 10]. In this case, $V_{IN} = \alpha \cdot \Delta V_{BE}$, while $V_{REF} = V_{BE} + \alpha \cdot \Delta V_{BE}$. A straightforward implementation would be

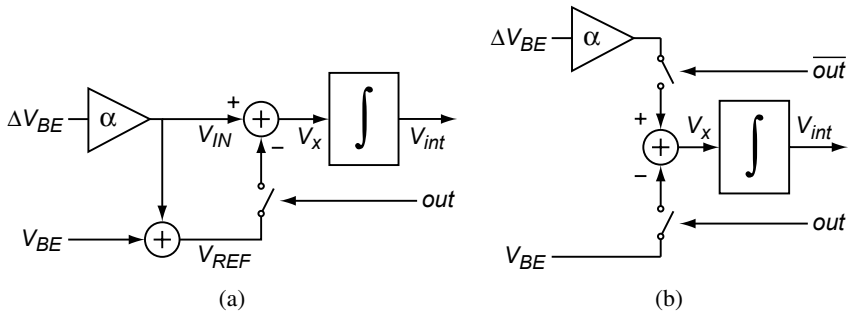


Figure 4.2. Realization of the transfer of a smart temperature sensor using charge balancing: (a) straightforward implementation, (b) optimized implementation.

to first generate V_{IN} and V_{REF} , and then apply them to the modulator (Figure 4.2a). The integrator's input voltage V_x is then:

$$V_x = \begin{cases} \alpha \cdot \Delta V_{BE} & \text{if } out = 0 \\ -V_{BE} & \text{if } out = 1 \end{cases} \quad (4.12)$$

This shows that exactly the same transfer can be obtained using the simpler configuration shown in Figure 4.2b. In this case, either $\alpha \cdot \Delta V_{BE}$ or $-V_{BE}$ is applied to the integrator. An explicit reference voltage V_{REF} is not needed anymore. The gain α , which is shown as a separate block, can now also be implemented in the integrator, by switching its gain to an α times larger value if $out = 0$.

4.1.4 Synchronous versus Asynchronous Modulation

The modulator of an indirect ADC can be either *synchronous* or *asynchronous*. Asynchronous modulators are also called free-running modulators. Duty-cycle modulators, period modulators, and frequency modulators are examples of asynchronous modulators. In these modulators, the final synchronization to a digital clock takes place in the conversion of the time-domain output of the modulator to a conversion result. In the case of a duty-cycle modulator, for instance, this conversion would typically be implemented by sampling the output of the modulator and counting the number of 1's in a given number of samples.

A synchronous modulator, in contrast, is synchronized to a clock, typically the same clock to which the output signal of the ADC is synchronized. Examples of synchronous indirect ADCs are dual-slope ADCs and $\Sigma\Delta$ ADCs. An example of a first-order $\Sigma\Delta$ modulator is shown in Figure 4.3. It is very similar to the duty-cycle modulator of Figure 4.1: only the comparator with hysteresis has been replaced by a *clocked* comparator without hysteresis. This

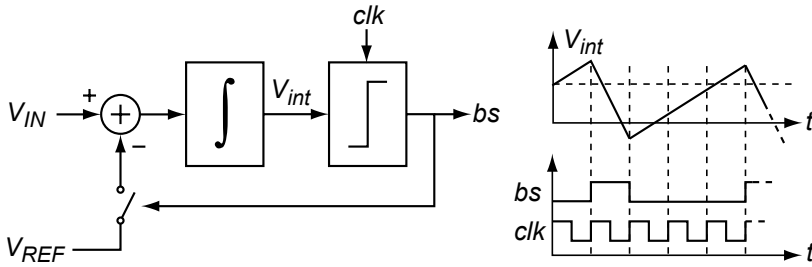


Figure 4.3. Block diagram and timing of a first-order $\Sigma\Delta$ modulator.

comparator changes its output on the rising edges of the clock only. Its output is a synchronous sequence of 0's and 1's, the *bitstream*.

The feedback in this modulator drives the output of the integrator back to zero. This form of charge balancing ensures that the average charge accumulated in the integrator is (approximately) zero (just as in a duty-cycle modulator). If, in a total number of clock cycles N_{total} , the bitstream is one during N_1 clock cycles, the charge balancing implies

$$N_{total} \cdot V_{IN} = N_1 \cdot V_{REF}, \quad (4.13)$$

which can be rewritten as

$$\frac{N_1}{N_{total}} = \frac{V_{IN}}{V_{REF}}. \quad (4.14)$$

This shows that the fraction of ones in the bitstream is equal to the ratio of V_{IN} and V_{REF} . This 'bit density' can be determined using a simple counter.

The synchronous and asynchronous approaches each have their pros and cons [7, 11]. The most important of these are the following:

- *Locking*: in the presence of periodic interference, asynchronous modulators have a tendency to lock to the frequency of the interferer if the modulator's frequency is close to (a fraction of) that frequency. Under such conditions, the linearity and resolution of the modulator can be severely reduced. Interference may result, for instance, from the clock of a digital circuit on the same chip as the modulator. Synchronous modulators are much less sensitive to locking. In some cases, the synchronization between a digital clock and the modulator can even be exploited to ensure that the digital clock edges occur at a moment that the modulator is least sensitive to them.
- *Normal-mode rejection*: synchronous modulators can be made more immune to interference present in their input signal (e.g. 50 Hz or 60 Hz interference from the power line) than asynchronous modulators. This is because

their conversion time can be made equal to an integer number of periods of the interfering signal. Dual-slope ADCs, for instance, often integrate their input signal during an integer number of power-line cycles. In a smart temperature sensor, where the inputs of the ADC are generated on-chip, this advantage is not very important, provided that the front-end circuitry has a sufficiently high AC power-supply rejection. Dynamic element matching generated in the front-end circuitry (such as those resulting from chopping or dynamic element matching) can be synchronized to the modulator (irrespective of whether the modulator itself is synchronous or asynchronous). Both modulator types can therefore be used to filter out these error signals (see also Section 4.6).

- *Resolution*: the resolution that can be obtained using the two modulator types in a given conversion time is roughly the same, if the modulators and their digital filters have the same order, and run at the same frequency (comparing, for instance, a duty-cycle modulator with a first-order $\Sigma\Delta$ modulator). In the case of an asynchronous modulator, the resolution can be increased by increasing the rate at which the output of the modulator is sampled. $\Sigma\Delta$ modulators, in contrast, typically run at the same clock frequency as the digital filter. The resolution obtained from a $\Sigma\Delta$ modulator can be increased either by increasing the clock frequency, or by increasing the order of the modulator and the digital filter. Higher-order asynchronous modulators (sometimes referred to as asynchronous $\Sigma\Delta$ modulators), in contrast, provide a wider bandwidth, but not necessarily a higher resolution [12].
- *Complexity*: the two modulator types are comparable in complexity if they have the same order. Higher-order modulators have a higher circuit and design complexity. One could say that synchronous modulators are somewhat more complex, as they require an extra oscillator to generate a clock signal. However, a clock signal is also needed for ADCs based on asynchronous modulators, in order to sample the modulator's output.

If the complexity and interface of a temperature sensor are to be kept simple, an asynchronous modulator may be preferable. A duty-cycle modulator, for instance, can then be integrated on the sensor chip, while a microcontroller samples its output and calculates the duty cycle to obtain a digital temperature reading [5]. A separate clock signal is then not needed on the sensor chip. The sensor outputs the intermediate time-domain signal of the converter, which only requires a single pin on the sensor's package. The required digital signal processing, such as counting and scaling to obtain a reading in degrees Celsius, is left to be implemented by the user.

Since the focus of this work is to design completely integrated smart temperature sensors, which provide a temperature reading in a readily interpretable

format, the digital processing has to be performed on-chip. In that case, an ADC based on a synchronous modulator is to be preferred, because of its relatively low sensitivity to interference and its compatibility with on-chip digital circuitry. $\Sigma\Delta$ ADCs are nowadays the most widely used indirect synchronous ADCs. An interesting advantage of these ADCs, compared to, for instance, dual slope converters, is the possibility to increase the resolution that can be obtained in a given number of clock cycles by using higher-order modulators. In the next section, the operating principles of $\Sigma\Delta$ ADCs will be discussed in more detail.

4.2 Operating Principles of Sigma-Delta ADCs

A first-order $\Sigma\Delta$ converter was already introduced in the previous section as an example of a synchronous indirect ADC. In this section, the operating principles of $\Sigma\Delta$ ADCs will be discussed in more detail, so as to provide background information for the discussion of first- and second-order $\Sigma\Delta$ modulators and their decimation filters later in this chapter.

4.2.1 Sampling and Quantization

For a discussion of $\Sigma\Delta$ ADCs, it is useful to take a brief look at the general operation of ADCs in the frequency domain [4, 13, 14]. Any ADC performs two basic functions on its input signal: quantization and sampling. Therefore, an ADC can be represented as a combination of a sample-and-hold circuit and a quantizer.

In the frequency domain, sampling results in a repetition of the input spectrum at multiples of the sampling frequency f_s , as illustrated by the dotted lines in Figure 4.4a. Overlapping of these replica is referred to as aliasing, and results in an irreversible distortion of the input signal. To prevent this, f_s has to be at least twice the bandwidth B of the input signal (the so-called Nyquist criterion), or, conversely, the bandwidth of the input signal has to be limited to $f_s/2$. This bandwidth limitation is usually performed by a so-called anti-aliasing filter which precedes the sample-and-hold circuit.

Quantization can be modelled as the addition of an error signal at the output of the ADC. If the input signal is sufficiently ‘busy’ (i.e. it changes from sample to sample), this error signal can be modelled as a uniformly distributed random variable, and is often referred to as quantization *noise*². Such noise has a white spectrum as indicated by the shaded area in Figure 4.4a.

²Note that a DC input signal, as found in a temperature sensor, is not a ‘busy’ signal; the implications of this will be discussed in Section 4.2.5.

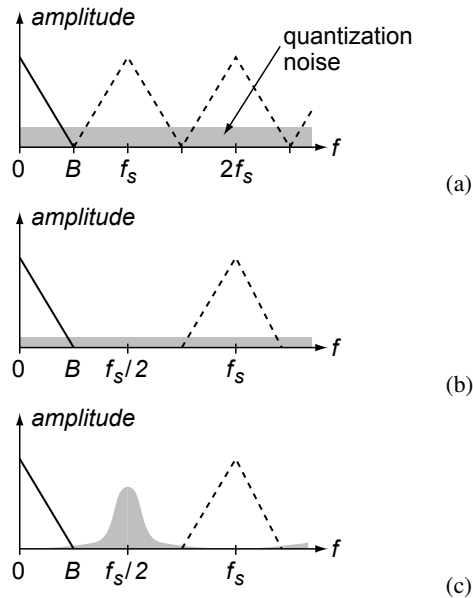


Figure 4.4. Signal and quantization noise spectrum for (a) a Nyquist ADC, (b) an oversampling ADC, and (c) a noise-shaping oversampling ADC.

4.2.2 Oversampling

ADCs operated at twice the bandwidth of the input signal are referred to as Nyquist ADCs, while ADCs that sample the input signal faster than required by the Nyquist criterion are called *oversampling* ADCs (Figure 4.4b). $\Sigma\Delta$ ADCs belong to this group, but in principle any ADC is an oversampling ADC if its conversion rate is higher than twice the signal bandwidth. The ratio $f_s/2B$ is called the oversampling ratio, and is 1 for a Nyquist ADC.

Oversampling ADCs have two main advantages compared to Nyquist ADCs. The first advantage is related to the anti-aliasing filter. For a Nyquist ADC, this filter should ideally be a brick-wall low-pass filter, which passes all signal components below $f_s/2$ and blocks everything above $f_s/2$. For an oversampling ADC, the requirements for the anti-aliasing filter can be relaxed, because the frequency band from B to $f_s/2$ is available for roll-off.

The second and more important advantage of oversampling ADCs lies in their ability to achieve a resolution higher than that of the quantizer. As the energy of the white quantization noise is spread over the frequency band from DC to f_s , the amount of noise in the signal band decreases if the sampling frequency is increased (4.4b). The quantization noise in the band B to $f_s/2$ can be filtered out by a digital filter, effectively reducing the quantization noise power by the oversampling ratio. For every doubling of the oversampling

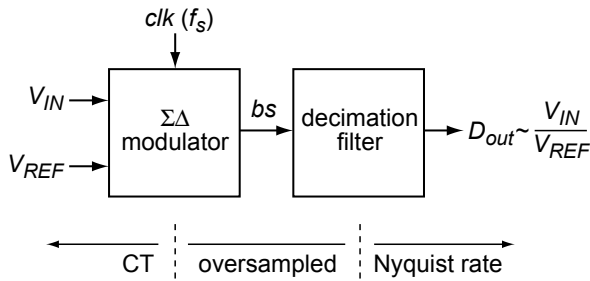


Figure 4.5. Block diagram of a $\Sigma\Delta$ ADC

ratio, the quantization noise power in the signal band is thus halved, effectively increasing the resolution by a 0.5 bits.

4.2.3 Noise Shaping

In a *noise-shaping* oversampling ADC, the spectrum of the quantization noise is shaped, so as to move even more of it outside the signal band, where it can be filtered out (Figure 4.4c). Thus, much more than 0.5 bits per doubling of the oversampling rate can be obtained, so that a high resolution can be obtained using a low-resolution quantizer. This property makes oversampling ADCs attractive for implementation in CMOS technology: resolution can be traded for speed. By using a comparator as a single-bit quantizer, the need for precise components (such as matched resistors or capacitors) is avoided. In return, the comparator needs to be fast, because of the oversampling, but this is usually not a problem in modern CMOS technology.

$\Sigma\Delta$ ADCs are the most common noise-shaping oversampling ADCs. A $\Sigma\Delta$ ADC consists of a $\Sigma\Delta$ modulator and a decimation filter (Figure 4.5). The modulator produces a bitstream bs , which is an oversampled single-bit representation of the ratio V_{IN}/V_{REF} . The decimation filter removes the shaped quantization noise from this bitstream, and brings the data rate back to the Nyquist frequency, producing the multi-bit output data D_{out} .

In the modulator, the noise shaping is achieved by incorporating the comparator in a feedback loop with a filter. In the case of the first-order modulator of Figure 4.3, this loop filter is formed by the integrator. In general, the loop filter can be a switched-capacitor (SC) filter or a continuous-time (CT) filter. In the case of a SC loop filter, the input and reference voltages are sampled directly at the input of the modulator (Figure 4.6a). In the case of CT loop filter, in contrast, the sampling takes place after the filter (Figure 4.6b). This has the advantage that the loop filter also acts as anti-aliasing filter.

In both modulator types, the reference of the modulator is multiplied by the bitstream (by means of a switch, which acts as a single-bit DAC) to obtain an

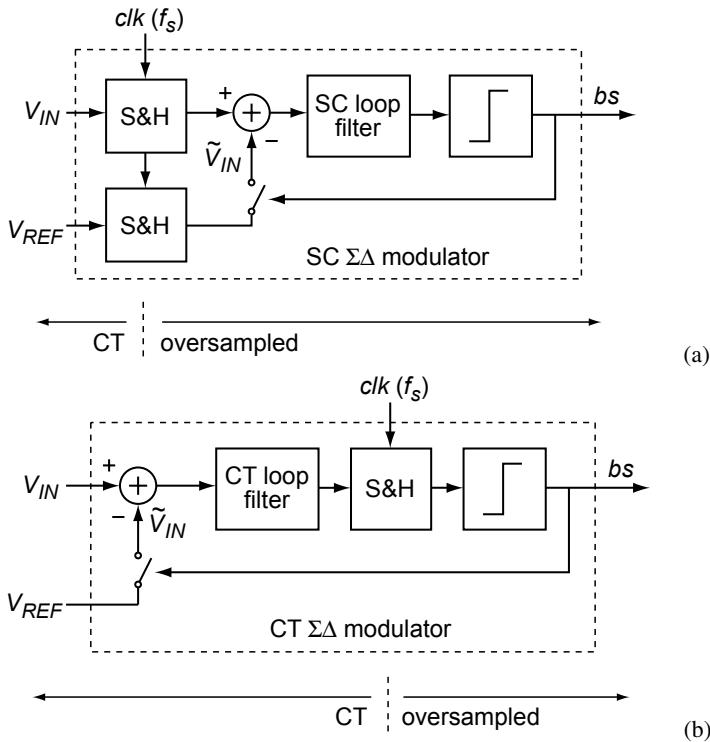


Figure 4.6. (a) Block diagram of a switched-capacitor (SC) $\Sigma\Delta$ modulator; (b) block diagram of a continuous-time (CT) $\Sigma\Delta$ modulator.

instantaneous approximation \tilde{V}_{in} of the input voltage. This approximation is subtracted from the actual input signal to obtain the instantaneous quantization error of the bitstream. This error is low-pass filtered by the loop filter, the output of which is thus a representation of the *average* quantization error of the bitstream. The comparator acts so as to drive this average error to zero, thus producing a bitstream whose average value equals the ratio V_{IN}/V_{REF} .

4.2.4 Linear Model

The behavior of $\Sigma\Delta$ modulators in the frequency domain can be understood best by considering the linear discrete-time model shown in Figure 4.7 [13]. In this model, the loop filter is modelled by its transfer function $G(z)$ (normalized to the reference voltage V_{REF}). The quantizer, which is the component that makes a $\Sigma\Delta$ modulator non-linear, is modelled as a gain k and an additive quantization noise source $E(n)$. For a multi-bit quantizer, the gain k is well-defined, but for a single-bit quantizer (a comparator), k has to be chosen so as to obtain a good match between simulation results obtained with the linear

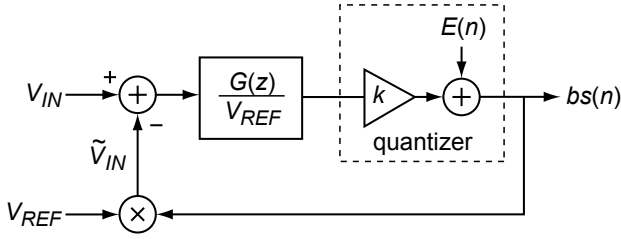


Figure 4.7. Discrete-time model of a $\Sigma\Delta$ modulator.

model and an actual modulator. The discrete-time model of Figure 4.7 fits best with SC modulators, which have a sampled loop filter. CT modulators can be modelled using a similar model preceded by a continuous-time pre-filter [15], so that the following also applies to these modulators.

The bitstream output bs of the modulator is a function of the input signal V_{IN}/V_{REF} , and the quantization noise $E(n)$:

$$bs(z) = H_{STF}(z) \frac{V_{IN}(z)}{V_{REF}} + H_{NTF}(z)E(z), \quad (4.15)$$

where $H_{STF}(z)$ is the signal transfer function:

$$H_{STF}(z) = \frac{kG(z)}{1 + kG(z)}, \quad (4.16)$$

and $H_{NTF}(z)$ is the noise transfer function:

$$H_{NTF}(z) = \frac{1}{1 + kG(z)}. \quad (4.17)$$

If $G(z)$ is a low-pass function, the signal transfer function is also a low-pass function, while the noise transfer function is a high-pass function. This shows that the quantization noise is high-pass filtered, and thus shaped to higher frequencies, conform Figure 4.4c. The higher the order of the loop filter, the more the quantization noise is shaped, and therefore the less quantization noise remains in the signal band. In-band quantization noise can therefore not only be reduced by increasing the oversampling ratio, but also by increasing the order of the loop filter.

Unfortunately, higher-order loop filters can be unstable. Moreover, they are more complex and require a higher-order, more complex decimation filter. It is therefore generally desirable to minimize the order of the loop filter for a given quantization noise requirement and oversampling ratio. The bandwidth in a smart temperature sensor is typically so low (10 Hz) that even with modest

clock frequencies (tens of kHz), high oversampling ratios can be obtained. As a result, first and second-order loop filters are sufficient for this application (as will be shown in more detail in the following sections).

4.2.5 Incremental Operation

So far, the input signal was assumed to be a ‘busy’ signal, so that the quantization error could be modelled as additive white noise. Unfortunately, the input signal in a temperature sensor, and many other instrumentation applications, is far from busy: it is essentially a DC signal. As a result, the quantization noise is not white, but concentrated at discrete frequencies (or ‘tones’), especially in first-order modulators [13]. Nevertheless, $\Sigma\Delta$ ADCs can be used successfully for this type of input signals, as will become clear in the following sections.

A first-order $\Sigma\Delta$ modulator for instrumentation applications was already introduced by Van der Plassche in 1978 [16]. The use of $\Sigma\Delta$ ADCs for instrumentation applications differs significantly from the typical use of such ADCs in audio or communication applications [17]. In the latter, the input signal is an AC signal, and the converter operates continuously, providing a stream of decimated output values. Its performance is typically expressed in terms of its dynamic range (DR) and its signal-to-noise and distortion ratio (SNDR). In instrumentation applications, in contrast, the input signal is a low-frequency or DC signal. The converter is typically operated in ‘single-shot’ mode, which means that it powers up, produces a single conversion result (e.g. a temperature reading) and finally powers down again to save power. The converter’s performance is measured in terms of its absolute accuracy (offset, gain and linearity), which will be expressed as the effective number of bits (ENOB) in this work (see Section 4.1.1).

$\Sigma\Delta$ converters tailored to instrumentation applications are usually referred to as ‘incremental ($\Sigma\Delta$) converters’ [17–19]. Their characteristics can be summarized as follows:

- The loop filter and the decimation filter are reset at the beginning of a conversion.
- The modulator does not operate continuously, but runs for a limited number of N clock cycles, producing a bitstream of N bits.
- The decimation filter is a finite impulse-response (FIR) filter with an response length of N .
- The decimation factor is also equal to N , so that the conversion result is thus a simple weighted sum of the bits of the bitstream.

The use of first- and second-order $\Sigma\Delta$ modulators and the associated decimation filters for such incremental converters will be discussed in the following sections.

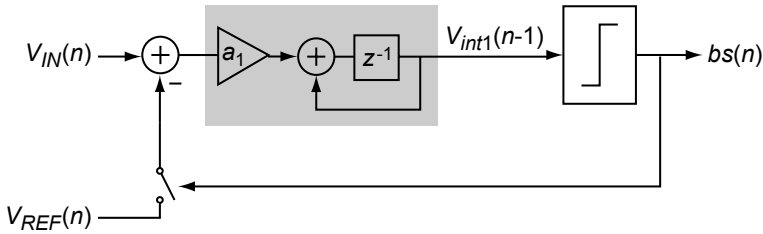


Figure 4.8. Discrete-time model of a first-order $\Sigma\Delta$ modulator.

4.3 First-Order Sigma-Delta Modulators

In this section, the characteristics and limitations of first-order $\Sigma\Delta$ modulators will be discussed. In the next section, the concepts introduced here will be extended to second-order modulators, which are more amenable to the higher resolution required in precision temperature sensors.

4.3.1 Topology

The block diagram of a first-order $\Sigma\Delta$ modulator has already been shown in Figure 4.3. A discrete-time model of this modulator is shown in Figure 4.8. Its loop filter is formed by a single integrator. The value of the coefficient a_1 , which determines the time-constant of the integrator, ideally does not affect the operation of the modulator. This is because the integrator is followed by a comparator, which only detects the sign of the integrator's output, so as to determine the bitstream value for the next clock cycle. In practice, a_1 should be low enough to prevent clipping of integrator's output.

4.3.2 Noise Shaping

According to the linear model discussed in the previous section, a first-order $\Sigma\Delta$ modulator should show first-order noise shaping, i.e. a quantization noise that increases by 20 dB/dec towards $f_s/2$. Figure 4.9 shows simulated power spectra of the bitstream of a first-order modulator, for both a DC input and a sinusoidal input signal. Clearly, the linear model is not accurate for the DC signal: the quantization noise is concentrated at discrete frequencies. The DC value of the bitstream, as will be shown below, is nevertheless still an accurate representation of the DC input signal. Therefore, this non-ideal noise shaping does not prevent the use of a first-order modulator for instrumentation purposes.

4.3.3 Resolution

An ideal first-order $\Sigma\Delta$ modulator produces a unique bitstream for every DC input [20]. However, if only the first N bits of the bitstream are used

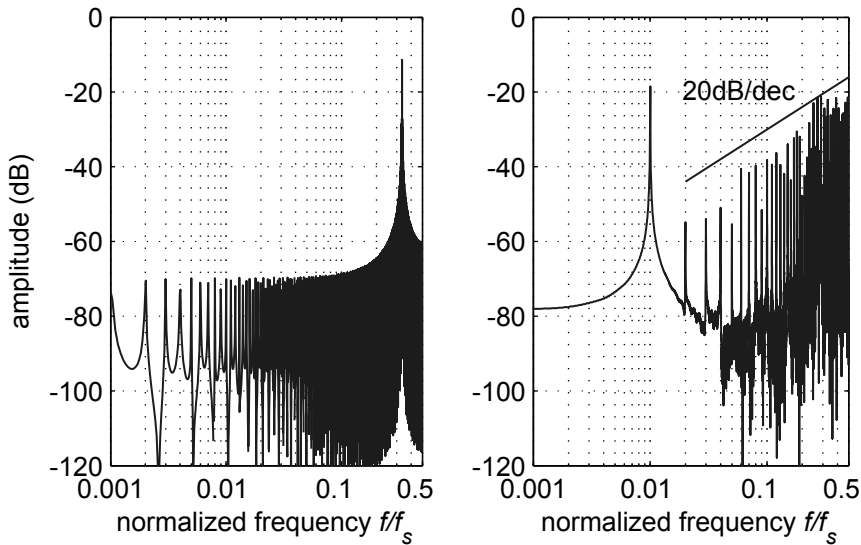


Figure 4.9. Spectrum of the bitstream (FFT of 16384 bits) of a first-order modulator with a DC input of $V_{IN} = 0.333V_{REF}$ (left) and a sinusoidal input at $f_s/100$ with an amplitude of $V_{REF}/4$ (right).

to produce a conversion result, the resolution is limited by the fact that the bitstreams corresponding to a range of DC inputs share the same first N bits. This is a fundamental limit of the modulator, unrelated to the decimation filter used to produce the conversion result. An ‘optimal’ decimation filter processes a bitstream in such a way that its output corresponds to the middle of the range of input values that share this bitstream. In theory, such a filter could be implemented as a lookup table with 2^N entries, which maps bitstreams to conversion results.

While such an ‘optimal’ decimation filter is impractical to implement, it can be simulated in order to find the best-case resolution of a $\Sigma\Delta$ modulator. The modulator’s input is swept from zero to V_{REF} in small steps, and the resulting sequences of N bits are recorded. Then, for every range of input values that results in the same sequence, a conversion result is produced that corresponds to the middle of that range. Finally, the quantization error is found as the difference between the conversion results and the input values.

The result of such a simulation of a first-order $\Sigma\Delta$ modulator is shown in Figure 4.10a for $N = 128$. The quantization error is largest when the ratio V_{IN}/V_{REF} is a rational number with a small denominator, for example around $V_{IN}/V_{REF} = 0, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}$ and 1. The peak error, which determines the ENOB, is 0.5 LSB of 7 bits.

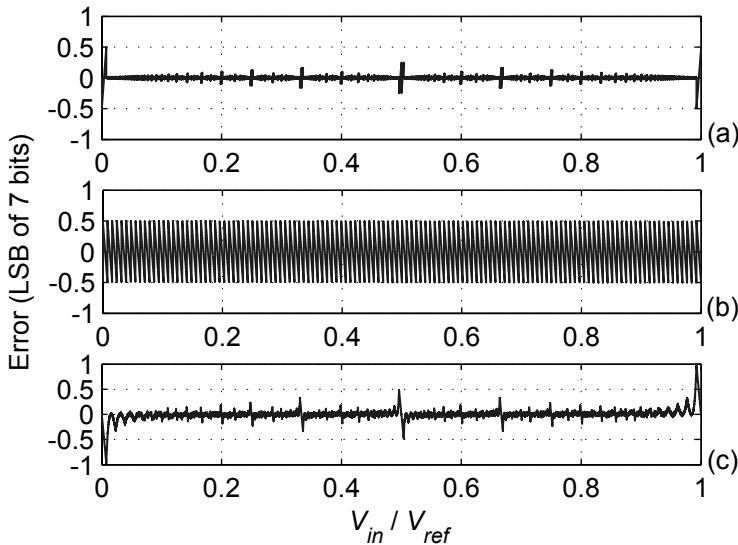


Figure 4.10. Quantization error of a first-order $\Sigma\Delta$ converter as a function of the DC input level, using $N = 128$ cycles, and (a) an optimal decimation filter, (b) a sinc filter, and (c) a sinc^2 filter.

The simulated ENOB as a function of the number of cycles N is shown by the bold line in Figure 4.11. As derived in Appendix B.1, this line can be described by

$$\text{ENOB}_{1st,ideal} = \log_2(N), \quad (4.18)$$

which shows that for every doubling of the number of cycles, one extra bit of resolution is obtained.

In practice, a decimation filter has to be used that can be implemented efficiently, and preferably does not significantly increase the peak quantization error compared to an optimal filter. A simple time-domain analysis shows that a sinc filter is a suitable decimation filter for a first-order modulator [18]. Such a filter has a rectangular impulse response, and can be implemented as a counter that determines the number of 1's in the first N bits of the bitstream. Figure 4.10b shows the error in the resulting conversion result. While for many input levels the quantization error is much larger than that obtained with an optimal filter, the peak error is the same, so that the ENOB is not reduced.

The use of a higher order decimation filter may locally increase the resolution compared to a sinc filter, but it will not improve the ENOB, since the sinc filter already reaches the optimal. Figure 4.10c, for instance, shows the quantization error obtained with a sinc^2 decimation filter. The peak error at the extremes of the input range is in this case even higher than for a sinc filter. If the regions

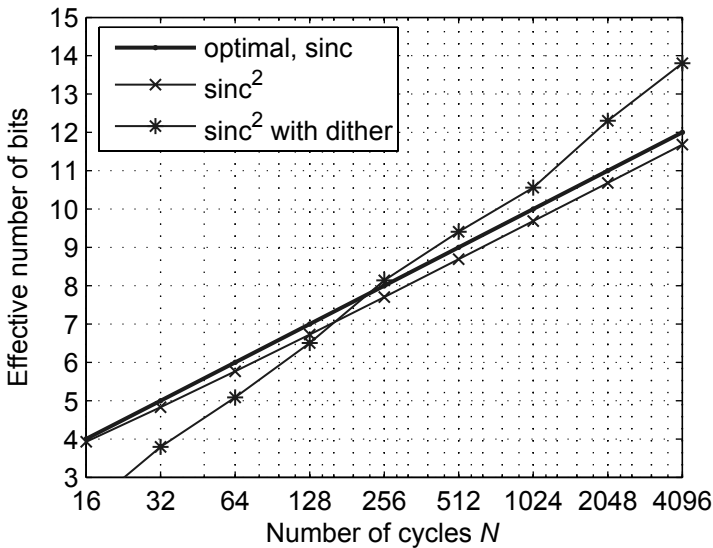


Figure 4.11. Maximum effective number of bits obtainable from a first-order $\Sigma\Delta$ modulator as a function of the number of cycles N , for various decimation filters.

around the extremes are avoided by restricting the input range, the ENOB is roughly equal to that obtained with a sinc filter (see Figure 4.11).

If a 15 bit resolution is required, as in a precision temperature sensor (see Section 4.1.1) $2^{15} = 32768$ clock cycles of a standard first-order $\Sigma\Delta$ modulator are needed. To obtain a conversion time of 100 ms, a clock frequency of about 325 kHz is needed. Such a high clock frequency is undesirable in terms of power consumption. In some cases, an extra bit of resolution can be obtained by using the fact that the output of the integrator at the end of the conversion is proportional to the quantization error [18]. The comparator can be used to determine the polarity of the quantization error, thus resulting in an extra bit of resolution. With this improvement, the required clock frequency is reduced to about 160 kHz. As this is still a fairly high frequency, further improvement is desired.

A way to improve the ENOB is to add dither to the modulator, either at its input, or at the input of the comparator [17]. In the latter case, the dither signal is attenuated by the gain of the integrator at low frequencies, and hence should not affect the in-band noise too much. Figure 4.11 shows the simulated ENOB if dither is added to the input of the comparator. The dither used in this simulation was white noise with a standard deviation of $a_1 V_{REF}/2$. The modulator now performs much more like an ideal modulator with a busy input signal [13]: the resolution increases by 1.5 bits for every doubling of the number of cycles. For

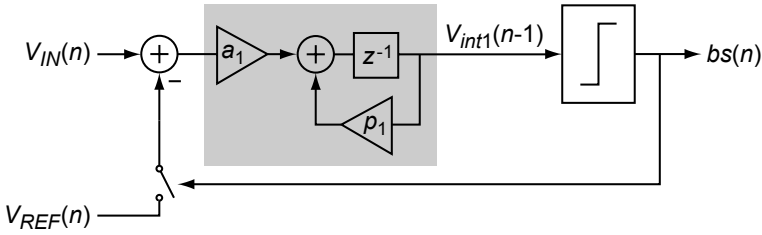


Figure 4.12. Discrete-time model of a first-order $\Sigma\Delta$ modulator with a leaky integrator.

$N \geq 256$, the ENOB is improved compared to that obtained from a sinc filter. It is also possible to use a pseudo-random dither signal, in which case the dither signal can be designed such that it is matched to the decimation filter of the $\Sigma\Delta$ modulator and hence completely filtered out [21].

With dither, roughly 6500 clock cycles are needed to obtain 15 bits, resulting in a clock frequency of 65 kHz. This improvement comes at the cost of a more complex sinc^2 decimation filter and the extra circuitry needed to generate the dither signal. As will be shown in Section 4.4, a second-order modulator can achieve 15 bits at a much lower clock frequency without dithering, while using the same sinc^2 decimation filter. The extra integrator required in a second-order modulator has relaxed specifications. Such a modulator is therefore more attractive for use in precision temperature sensors.

4.3.4 Leakage

In a practical implementation, the integrator of a first-order $\Sigma\Delta$ modulator will have a finite DC gain A_0 . This results in leakage: the charge on the integration capacitor slowly leaks away, even if the integrator's input is zero. Figure 4.12 shows how this effect can be included in the model of the modulator [20]: a gain $p_1 < 1$ is incorporated in the feedback of the integrator, where

$$p_1 = 1 - \frac{a_1}{A_0}. \quad (4.19)$$

For a leakage-free integrator, with infinite DC gain, $p_1 = 1$.

In $\Sigma\Delta$ modulators, leakage results in non-linearity [20], and hence limits the achievable ENOB. While a modulator with a non-leaky integrator produces a unique bitstream for every DC input, this is not true for a modulator with a leaky integrator. The leakage causes the modulator to produce limit cycles (periodic sequences of bits) that are not associated with a unique input value, but rather with a range of input values. Input values that cause the modulator to lock into the same limit cycle form a 'dead zone', and will inevitably lead to the same conversion result, no matter how many bits of the bitstream are taken into

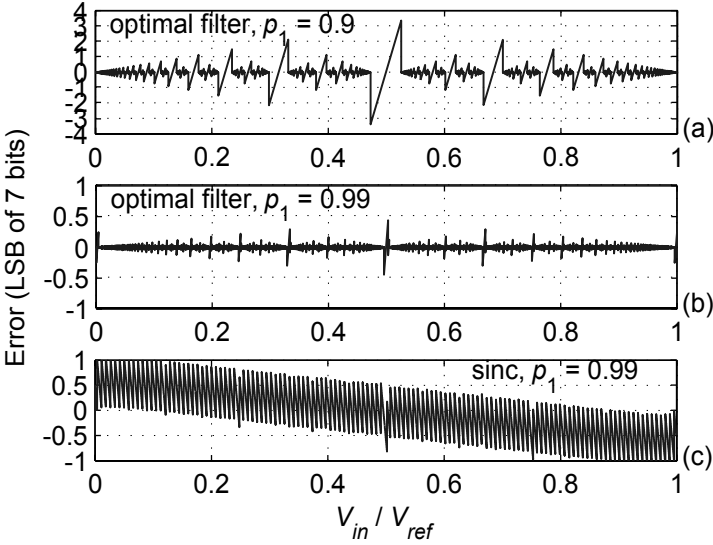


Figure 4.13. Quantization error of a leaky first-order $\Sigma\Delta$ converter as a function of the DC input level, using $N = 128$ cycles, and (a) an ideal decimation filter and 10% leakage, (b) an ideal decimation filter and 1% leakage, and (c) a sinc decimation filter and 1% leakage.

account. This locking behavior results in an input-output characteristic with a staircase shape, referred to as the ‘devil’s staircase’.

Figures 4.13a and b shows the simulated quantization errors of a first-order modulator with $p_1 = 0.9$ and 0.99 , respectively. These results were obtained using $N = 128$ cycles, and an optimal decimation filter (which, as before, maps a bitstream to the value that correspond to the middle of the associated input range). Clearly, for $p_1 = 0.9$, which corresponds to a DC gain in the order of 10, the peak quantization error is much larger than in the leakage-free case (see Figure 4.10a). For $p_1 = 0.99$, which corresponds to a ten times larger DC gain, the increase in error has become negligible.

Figure 4.14 shows the simulated ENOB as a function of N for various values of p_1 . If N is small, the ENOB is (approximately) equal to the ideal value given by (4.18). For larger values of N , the resolution becomes leakage-limited.

The maximum ENOB that can be obtained for a given amount of leakage is determined by the width of the largest dead zone. It can be shown (see Appendix B.1) that this width can be approximated by [20]:

$$\Delta V_{IN} = \frac{1 - p_1}{1 + p_1} V_{REF} \tag{4.20}$$

$$\simeq \frac{1 - p_1}{2} V_{REF} = \frac{a_1}{2A_0} V_{REF} \quad (A_0 \gg a_1). \tag{4.21}$$

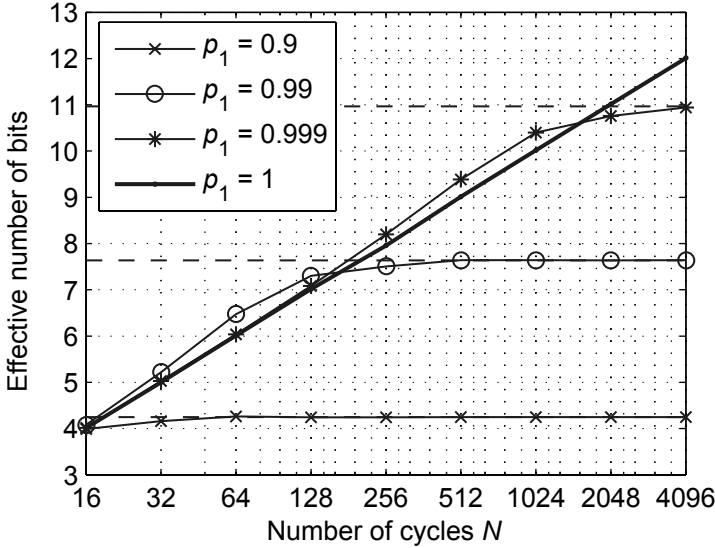


Figure 4.14. Maximum effective number of bits obtainable from a first-order $\Sigma\Delta$ modulator as a function of the number of cycles N , for different values of the integrator leakage p_1 ; the horizontal dashed lines indicate the limit given by equation (4.22).

An optimal decimation filter will map the limit cycle associated with this largest step to a conversion result that corresponds to the middle of the step. The peak quantization error is then equal to half the step width, and the effective number of bits is therefore bounded as follows:

$$\text{ENOB}_{1st,leaky} \leq \log_2 \left(\frac{V_{REF}}{\Delta V_{IN}} \right) \simeq \log_2 \left(\frac{A_0}{a_1} \right) + 1 \quad (A_0 \gg a_1). \quad (4.22)$$

This limit is indicated by the horizontal dashed lines in Figure 4.14 and agrees well with the simulation results.

Integrator leakage not only limits the linearity of the modulator. Just as in linear systems, the finite loop gain associated with leakage also results in a gain error: the average value of the bitstream output is not exactly equal to that of the input [20]. This gain error is compensated for by an optimal decimation filter, but with a practical decimation filter, such as a sinc filter, it is revealed (see Figure 4.13c). The quantization error shows the following trend:

$$D_{out} - \frac{V_{IN}}{V_{REF}} = -\varepsilon_A \left(\frac{V_{IN}}{V_{REF}} - \frac{1}{2} \right), \quad (4.23)$$

where the gain error ε_A is given by

$$\varepsilon_A = 1 - p_1 = \frac{a_1}{A_0}. \quad (4.24)$$

The decimation filter could be modified to compensate for this gain error. In practice, however, the gain error will spread. It is therefore preferable to design the integrator's DC gain such that the gain error is negligible. The required DC gain can be calculated using (4.6), which specifies the maximum gain error for a given maximum temperature error ΔT_{max} . For $\Delta T_{max} = \pm 0.01^\circ\text{C}$, for example, the gain error has to be smaller than $\pm 0.0025\%$, which implies a DC gain in the order of 92 dB.

4.3.5 Initialization

In the previous section, it was mentioned that incremental operation of a $\Sigma\Delta$ converter implies that the loop filter and the decimation filter are reset at the beginning of a conversion. This reset brings the integrator in a well-defined state, so that conversions with the same input result in the same output, as shown in Figure 4.15a. Obviously, this means that averaging of successive conversion results does not result in a higher resolution. If the possibility to improve the resolution by means of averaging is desired, two alternatives can be considered.

The first alternative is to leave the initial state of the integrator undefined. This introduces extra uncertainty in the decimation result. This is illustrated in Figure 4.15b, which shows the quantization error for successive conversions where the initial value of the integrator's output is random with a uniform distribution between $-a_1 V_{REF}$ and $+a_1 V_{REF}$ (which is equal to the output swing of the integrator under normal operation). The quantization error randomly jumps between two values 1 LSB apart, which indicates that the least significant bit is now random. Hence, the resolution obtained from a single conversion has decreased by one bit. Some improvement, however, can now be obtained by averaging of successive conversion results.

A second alternative to resetting the integrator is to preserve the state reached at the end of the previous conversion. Figure 4.15c shows the resulting quantization error of successive conversions, which reveals the 'tonal' behavior that makes first-order modulators unattractive for audio applications: a repetitive error pattern appears that depends on the DC input level and, in an audio application, would result in an audible tone. While this tonal behavior can be prevented by resetting the integrator at the start of every conversion, it can also be exploited to allow further increase of resolution by averaging of successive conversion results. Since this averaging effectively extends the length of the decimation filter, the resolution increases by one bit for every doubling of the number of conversions that is averaged.

This type of operation requires that the time between the conversions is short enough to prevent that the state of the integrator dissipates away as a result of leakage. This is, for instance, the case if the converter is operated continuously. Clearly, this type of operation cannot be used if the circuitry is powered down between conversions, as is often the case in temperature sensors.

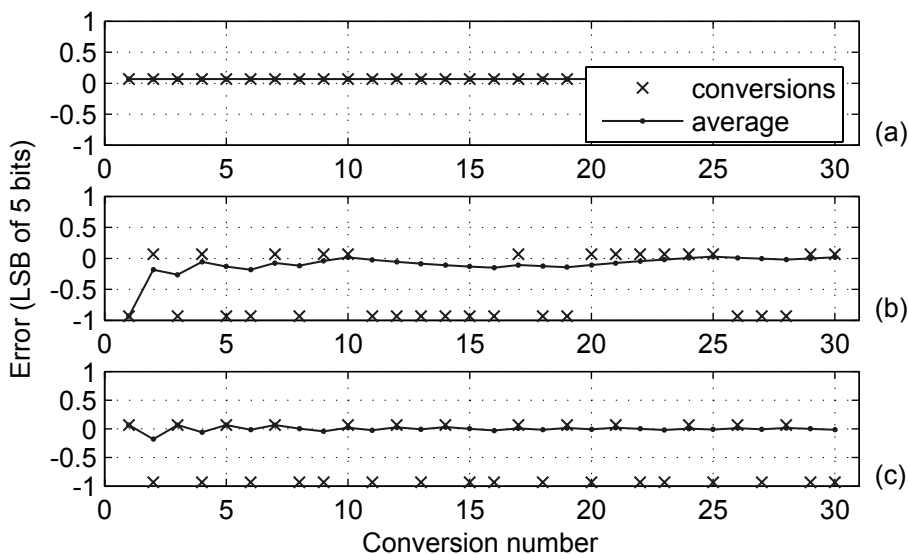


Figure 4.15. Quantization error and its running average for successive conversions of a first-order modulator with a sinc filter and $N = 32$, (a) with integrator reset, (b) with random initial integrator state, and (c) with preserved integrator state.

4.4 Second-Order Sigma-Delta Modulators

Second-order $\Sigma\Delta$ modulators can be used to obtain the resolution required in precision temperature sensors using much less clock cycles than first-order modulators. For a given conversion time, they thus allow a significant reduction in clock frequency and an associated reduction in power consumption. In this section, the characteristics and limitations of such modulators will be discussed.

4.4.1 Cascading versus Higher-Order Loop Filters

There are various ways in which the resolution that can be obtained in a given number of clock cycles can be improved compared to a first-order modulator: a second-order loop filter can be used, two first-order modulators can be cascaded, or multi-bit feedback can be applied [13]. The latter option is unattractive, since it requires accurate matching in the feedback DAC. Moreover, it is not very compatible with the simple charge-balancing scheme discussed earlier (see Figure 4.2). The other alternatives, a higher-order loop filter and cascading, deserve some more attention.

If a second-order loop filter is used, the quantization noise is shaped by 40dB/dec rather than by 20dB/dec [22]. As a result, less quantization noise remains in the signal band, which gives one bit of extra resolution per doubling of the oversampling ratio compared to a first-order modulator. A disadvantage of a second-order loop filter is that the resulting modulator is only stable for a restricted DC input range. For inputs close to zero or V_{REF} , the integrator outputs become excessively large and the modulator produces low-frequency limit cycles that result in a strong increase of the quantization error.

A cascade of two first-order modulators does not have this disadvantage. In such a cascade, the quantization error of a first modulator is processed by a second modulator. The outputs of the two modulators are digitally combined to obtain the conversion result [13]. Thus, a similar improvement in resolution can be obtained as for a modulator with a second-order loop filter. An incremental converter using such a cascaded topology was described in [19]. As a result of the unconditional stability of first-order modulators, a high resolution can be obtained over the full input range.

Whether the limited input range of a single-loop second-order modulator is a problem, depends on the charge-balancing configuration used. For the simple balancing of ΔV_{BE} against V_{BE} (see Figure 4.2b), only the middle third of the input range is used (see Figure 3.3), so that the regions with increased quantization noise are conveniently avoided. For the ‘more efficient’ configuration shown in Figure 3.3, about 90% of the input range is used, which is more problematic. For this configuration, a cascaded modulator would be more attractive.

A disadvantage of cascaded modulators is that they are sensitive to component mismatches. More specifically, the gain of the integrator of the first modulator has to be accurately defined to allow the second modulator to accurately quantize the error at its output. In a switched-capacitor implementation, this gain is typically defined by a ratio of capacitors and, with accurate layout, can therefore be accurate to about $\pm 0.1\%$. It can be shown that the resulting quantization errors are then negligible if an overall resolution of 16 bits is required [19]. In a continuous-time implementation, however, the integrator gain is defined by a resistance, the integration capacitance, and the clock period. Given the large spread on the absolute values of resistors and capacitors, the gain can easily spread by more than $\pm 10\%$ (unless special care is taken to derive the clock from a matched RC time constant). This makes cascaded modulators much less attractive for continuous-time implementation than single-loop modulators.

A further disadvantage of a cascaded modulator is that it requires an extra comparator and a more complex decimation filter. Given these considerations, the rest of this section will focus on single-loop second-order modulators.

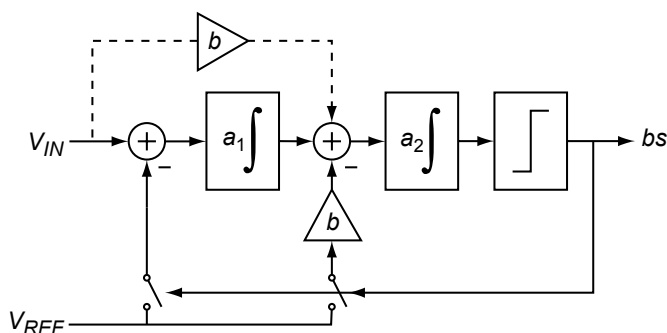


Figure 4.16. Block diagram of a second-order $\Sigma\Delta$ modulator. Inclusion of the feed-forward path (dashed) reduces the output swing of the first integrator.

4.4.2 Topology

Figure 4.16 shows a block diagram of a $\Sigma\Delta$ modulator with a second-order loop filter [22, 23]. The loop filter consists of two integrators with gains a_1 and a_2 , which can be modelled the same way as the integrator of the first-order modulator discussed in the previous section (see Figure 4.8).

To maintain stability, a feedback path with gain b to the input of the second integrator is used. The relative gain of this feedback path with respect to the gain of the first integrator, b/a_1 , determines how stable the loop is. If $b \gg a_1$, the effect of the first integrator is negligible, and the loop behaves as a first-order modulator, which is stable over the full input range. If, in contrast, $b \ll a_1$, the loop is unstable, resulting in very large integrator outputs. For intermediate values (a typical choice is $b/a_1 = 2$), the loop is conditionally stable: there is a range of input values for which the integrator outputs are bounded.

Since the second integrator is followed by a comparator, its gain a_2 ideally does not affect the operation of the modulator. In practice, the gains a_1 and a_2 are designed to prevent clipping at the integrator outputs.

When the modulator is used with a DC input, it is useful to include a feed-forward path from V_{IN} to the second integrator [24]. Without such a feed-forward path, the average output of the first integrator has to be bV_{IN} to ensure that the average input to the second integrator is zero. With a feed-forward path with gain b , the average output of the first integrator can remain zero. Thus, the peak-to-peak output swing of the first integrator is significantly reduced. It can be shown that the feed-forward path does not alter the noise transfer function of the modulator.

An additional advantage of the feed-forward path is that it simplifies the implementation. A feedback path without feed-forward path would require a separate reference voltage, which is not available in the optimized charge-

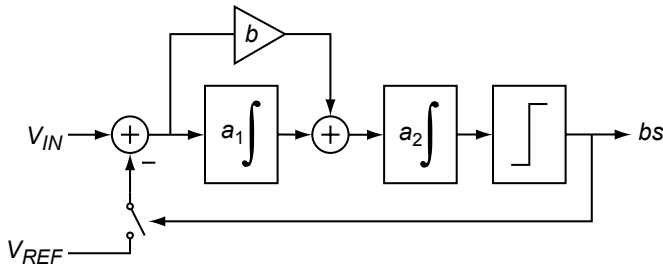


Figure 4.17. Simplification of the modulator of Figure 4.16.

balancing scheme of Figure 4.2b. By including the feed-forward path, the charge balancing between V_{IN} and V_{REF} at the input of the second integrator is identical to that at the input of the first integrator (except for the gain b). Therefore, the modulator can be redrawn as shown in Figure 4.17, which is compatible with the charge-balancing scheme of Figure 4.2b.

4.4.3 Stability

The output swing of both integrators as a function of the DC input level is shown in Figure 4.18. The curves shown are the maxima and minima of the integrator output voltages for different values of the ratio b/a_1 . They have been normalized to the reference voltage multiplied by the gains of the integrators. Since these gains are typically smaller than one, the voltage swing is typically smaller than the numbers on the graphs may suggest.

The ratio b/a_1 determines for what range of input values the integrator outputs are bounded. For b/a_1 smaller than 1.5, the outputs are not bounded for any input value, and the modulator can be considered unstable. For larger values of b/a_1 , the output swing of the first integrator is bounded for the full input range, and is almost independent of b/a_1 . The output of the second integrator, in contrast, blows up for input levels close the extremes of the input range. In practice, the output will clip, resulting in a strong increase in the quantization error. As mentioned before, this limits the usable input range.

How much of the input range of the modulator can be used, depends on the choice of the gains a_1 , a_2 and b , and on the level at which the integrator outputs start clipping. This is illustrated in Figure 4.19, which shows the percentage of the input range that is clipping-free as a function of the ratio b/a_1 , for three different clipping levels. For $b/a_1 < 1.5$, the outputs of the integrators clip due to instability. For large values of b/a_1 , the output of the second integrator clips due to the large signals from the feed-forward path. Irrespective of the clipping level, an optimum is found around $b/a_1 = 2$. This same value is used in [23].

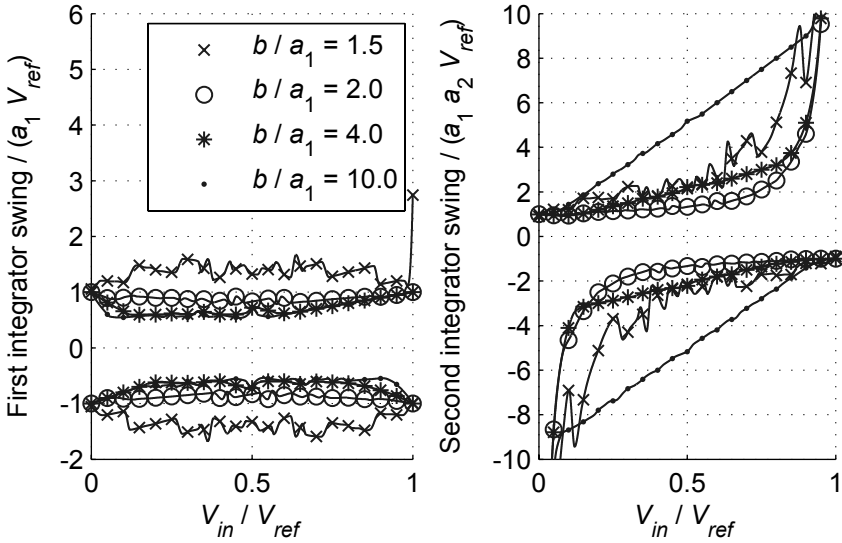


Figure 4.18. Simulated peak values (positive and negative) of the integrator outputs as a function of the DC input level, for different values of the feedback coefficient b .

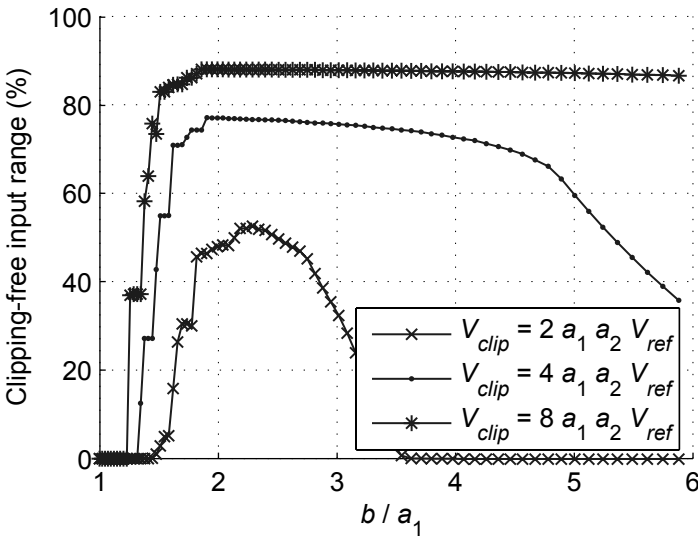


Figure 4.19. Fraction of the input range for which $|V_{int2}| < V_{clip}$, as a function of the feedback coefficient b .

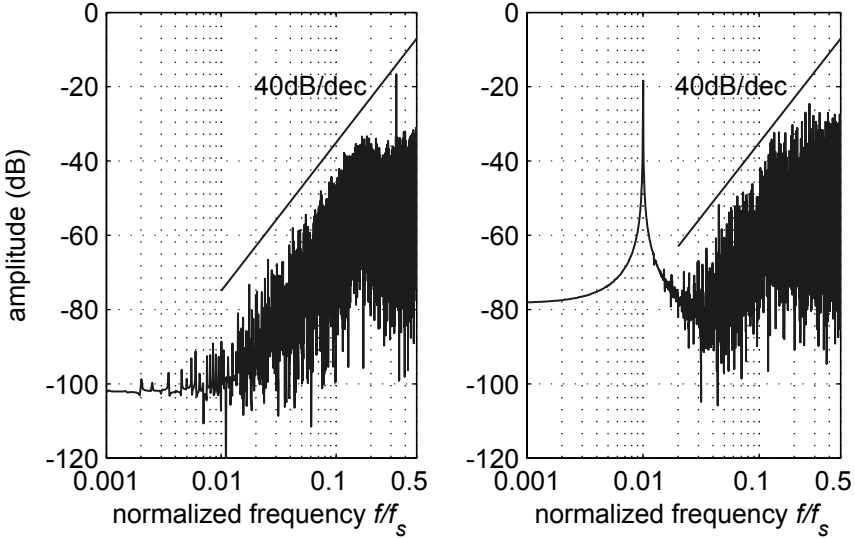


Figure 4.20. Spectrum of the bitstream (FFT of 16384 bits) of a second-order modulator ($b/a_1 = 2$) with a DC input of $V_{IN} = 0.333 \cdot V_{REF}$ (left) and a sinusoidal input at $f_s/100$ with an amplitude of $V_{REF}/4$ (right).

4.4.4 Noise Shaping

Figure 4.20 shows a simulated power spectrum of the bitstream of a second-order modulator with $b/a_1 = 2$, for both a DC input and a sinusoidal input signal. Clearly, the tonal behavior for a DC input is much less pronounced than that of a first-order modulator (see Figure 4.9). Both spectra clearly show second-order noise shaping.

4.4.5 Resolution

As for a first-order $\Sigma\Delta$ converter, the resolution of a second-order $\Sigma\Delta$ converter is limited by the fact that, for a given number of cycles N , there are ranges of input values that result in the same bitstream. This limit is independent of the decimation filter used. Decimation filters for use with second-order modulators will be discussed in Section 4.5. Here, the best-case ENOB will be evaluated using an ‘optimal’ decimation filter, which maps every bitstream of N bits to an output that corresponds to the middle of the range of inputs that give rise to that bitstream.

The simulated ENOB obtained using an optimal decimation filter is shown in Figure 4.21 for different values of b/a_1 . In this simulation, the peak quantization error was evaluated over a restricted input range from $0.3V_{REF}$ to $0.7V_{REF}$, which corresponds (with some margin) to the range used in a temperature sensor.

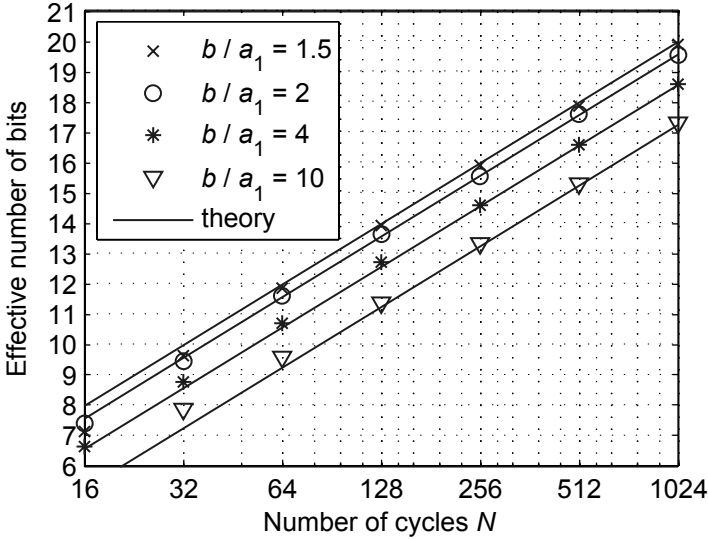


Figure 4.21. Simulated effective number of bits obtainable from a second-order $\Sigma\Delta$ modulator as a function of the number of cycles N , for different values of b/a_1 . The input signal was restricted to the range from 0.3 to 0.7. The lines indicate the values predicted by (4.25).

For $N \geq 128$, the ENOB increases by 2 bits every time the number of cycles is doubled. The ENOB is smaller for larger values of b/a_1 , which can be explained by the fact that a larger value of b/a_1 makes the noise shaping less aggressive (and, in consequence, the modulator more stable). The value $b/a_1 = 2$, which, as mentioned before, maximizes the usable input range, results in a near-optimal ENOB, and is therefore a good design choice.

As derived in Appendix B.2, the ENOB can be described by

$$\text{ENOB}_{2nd,ideal} = 2 \log_2(N) - \log_2 \left(\frac{2b}{3a_1} \right). \quad (4.25)$$

The ENOB predicted by this equation is indicated by the solid lines in Figure 4.21, and agrees well with the simulation results for $N \geq 128$. To obtain an ENOB of 15 bits, as required in a precision temperature sensor, from a modulator with $b/a_1 = 2$ combined with an optimal decimation filter, 209 cycles are required. It should be noted that practical decimation filters will require a larger number of cycles, as will be shown in Section 4.5, but still much less than the 32768 cycles required by a first-order modulator.

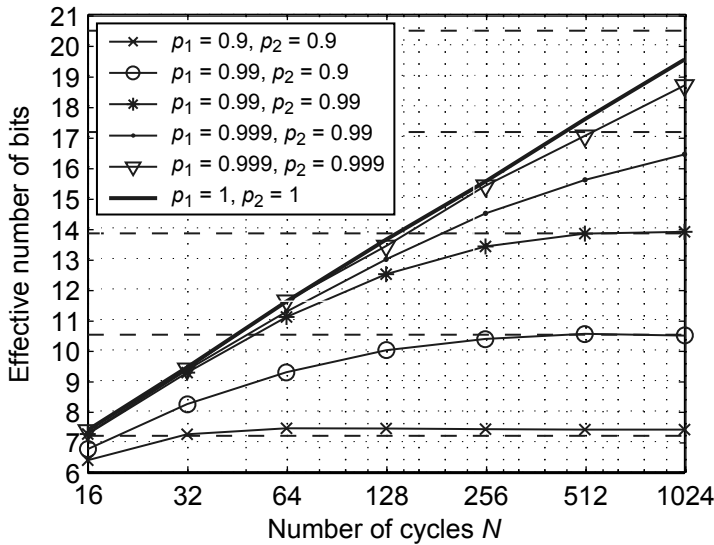


Figure 4.22. Maximum effective number of bits obtainable from a second-order $\Sigma\Delta$ modulator with $b/a_1 = 2$ as a function of the number of cycles N , for different values of the integrator leakages p_1 and p_2 ; the horizontal dashed lines indicate the limit given by equation (4.28). The input range was restricted from 0.3 to 0.7.

4.4.6 Leakage

As in a first-order modulator, integrator leakage limits the ENOB that can be obtained with a second-order modulator, and introduces a gain error. Leakage can be included in a discrete-time model of the modulator by including feedback coefficients p_1 and p_2 that are smaller than one, in a similar way as was done for the first-order modulator in Figure 4.12. These coefficients are related to the DC gains of the integrator as follows:

$$p_1 = 1 - \frac{a_1}{A_{0,1}}, \quad (4.26)$$

$$p_2 = 1 - \frac{a_2}{A_{0,2}}, \quad (4.27)$$

where $A_{0,1}$ and $A_{0,2}$ are the DC gain of the first and second integrator, respectively.

Figure 4.22 shows the simulated ENOB obtained using an optimal decimation filter for various values of p_1 and p_2 . The behavior is similar to that of a first-order modulator: for small values of N , the ENOB follows the ideal behavior expressed by (4.25), while for larger values it is limited by leakage.

The maximum ENOB that can be obtained from a leaky second-order modulator is determined by the width of the largest dead zone, and can be approximated by (see Appendix B.2):

$$\begin{aligned} \text{ENOB}_{2nd,leaky} &\leq \log_2 \left(\frac{V_{REF}}{\Delta V_{IN}} \right) \simeq \log_2 \left(\frac{3a_1}{b(1-p_1)(1-p_2)} \right) \\ &= \log_2 \left(\frac{3A_{0,1}A_{0,2}}{ba_2} \right) \quad (A_{0,1} \gg a_1, A_{0,2} \gg a_2). \end{aligned} \quad (4.28)$$

This limit is indicated by the dashed lines in Figure 4.22, and agrees well with the simulation results, except for $p_1 = p_2 = 0.9$, in which case higher-order terms, which have been omitted in (4.28), are not negligible.

Equation (4.28) shows that the ENOB limit in a second-order modulator is determined by the combined DC gain of the two integrators. To obtain an ENOB of 15 bits, a combined DC gain in the order of 90 dB is required (depending on the values of b and a_2). Usually, as will be explained in the next chapter, a relatively high DC gain will be needed in the first integrator to ensure accurate charge balancing, and to suppress errors and noise introduced by the second integrator. This means that the effects of leakage can be reduced to negligible levels even with a modest gain in the second integrator.

Leakage not only limits the resolution, but it also introduces a gain error: the DC value of the bitstream differs from the modulator's input. As a result, the quantization error obtained with a practical decimation filter show the following trend:

$$D_{out} - \frac{V_{IN}}{V_{REF}} = -\varepsilon_A \left(\frac{V_{IN}}{V_{REF}} - \frac{1}{2} \right), \quad (4.29)$$

where the gain error ε_A can be described by

$$\varepsilon_A = (1-p_1)(1-p_2) \frac{(b-a_1/2)}{a_1 + b(1-p_1)} \quad (4.30)$$

$$\simeq \frac{a_1 a_2}{A_{0,1} A_{0,2}} \left(\frac{b}{a_1} - \frac{1}{2} \right). \quad (4.31)$$

For $a_1 = 0$ (which reduces the second-order modulator to a first-order modulator) this error simplifies to (4.24). The required combined DC gain $A_{0,1}A_{0,2}$ to reduce this gain error to negligible levels can be calculated using (4.6), which specifies the maximum gain error for a given maximum temperature error ΔT_{max} . For $\Delta T_{max} = \pm 0.01$ °C, for example, the gain error has to smaller than $\pm 0.0025\%$, which implies a combined DC gain in the order of 96 dB.

4.5 Decimation Filters

A decimation filter for an incremental $\Sigma\Delta$ modulator should ideally filter out all quantization noise from the bitstream, while preserving its DC value. Since

only N bits of the bitstream are available, a practical decimation filter must have a finite impulse response (FIR) of length N . In this section, the design of such a FIR decimation filter will be discussed. It will also be shown how such a filter can be designed to perform linear scaling of the conversion result, and to compensate for non-linearity.

4.5.1 Filters Matched to the Loop Filter

In [19], Robert and Deval have described a decimation filter for a cascaded second-order incremental $\Sigma\Delta$ converter. Their filter has an impulse response that is matched to the impulse response of the loop filter of the modulator. It has a very simple implementation, and achieves an ENOB close to the optimal value. This technique has been extended to non-cascaded and higher-order modulators by Lyden [25, 26] and by Márkus [17].

The idea behind their approach is as follows. At the end of a conversion, the output of the loop filter equals its response to the DC input signal, minus its response to the bitstream. Since the feedback loop nulls the output of the loop filter, the loop filter's response to the bitstream must be proportional to the input signal. Hence, if the bitstream is passed through a decimation filter that is synchronized to the loop filter and has the same impulse response, the response of that filter at the end of a conversion is also proportional to the input signal, as desired. Since such a 'matched' decimation filter is synchronized to the loop filter, no startup cycles are needed to wait for the modulator to reach its steady state.

For the second-order $\Sigma\Delta$ modulator introduced in the previous section, a matched decimation filter consists of a cascade of two digital integrators with a feed-forward path from the input to the second integrator. The impulse response of such a filter has a non-symmetrical triangular shape, and is shown in Figure 4.23. It can be implemented very easily using a down-counter and an accumulator which adds the value of the counter whenever the bitstream is one. Figure 4.23 also shows, for comparison, a symmetrical triangular impulse response (which corresponds to a sinc^2 frequency response).

An important disadvantage of a matched decimation filter, however, is that it has a poor normal-mode rejection [17]. Figure 4.23 shows the simulated frequency response of the matched filter, and that of the triangular filter. The matched filter, while providing some low-pass filtering, does not have the notches of the triangular filter. Such notches are required if the $\Sigma\Delta$ ADC is used to completely filter out dynamic error signals from its input signal, such as the errors signals resulting from chopping or dynamic element matching in the front-end circuitry (see Section 4.6). In such a situation, matched filters are not very attractive.

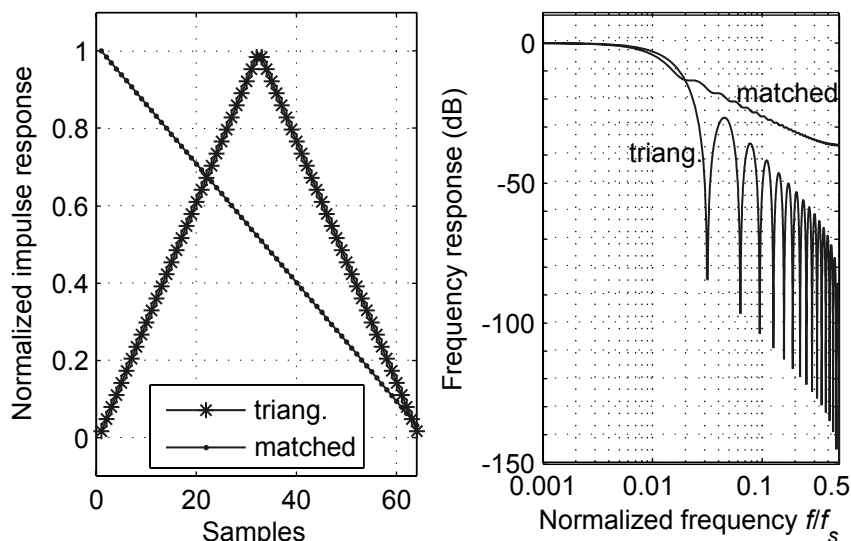


Figure 4.23. Impulse response (left) and frequency response (right) of a triangular window and a window that is matched to the loop filter.

4.5.2 Filters Based on Window Functions

Instead of using a filter matched to the modulator's loop filter, a decimation filter can also be designed by truncating the impulse response of an ideal filter using a window function. With an appropriate window function, filters can be found that result in a good ENOB performance, and also have the notches needed for filtering of dynamic error signals.

An ideal filter would be an infinite averaging filter. Such a filter has a frequency response that is zero at all frequencies except at DC, and an impulse response that is one at all times. A FIR filter with an impulse response of length N can be found by truncating this ideal impulse response by means of a window function [27]. This results in a decimation filter whose impulse response equals the window function. The choice of the window function then determines how well the shaped quantization noise of the modulator is suppressed.

There are many commonly used window functions. They are usually distinguished by the width of their pass band (the 'main lobe' of the filter) and the attenuation of the side lobes. For a given impulse response length N , the attenuation of the side lobes can only be reduced at the expense of an increase in the width of the main lobe [27]. Therefore, it is not obvious which window function will result in the highest resolution in combination with a second-order

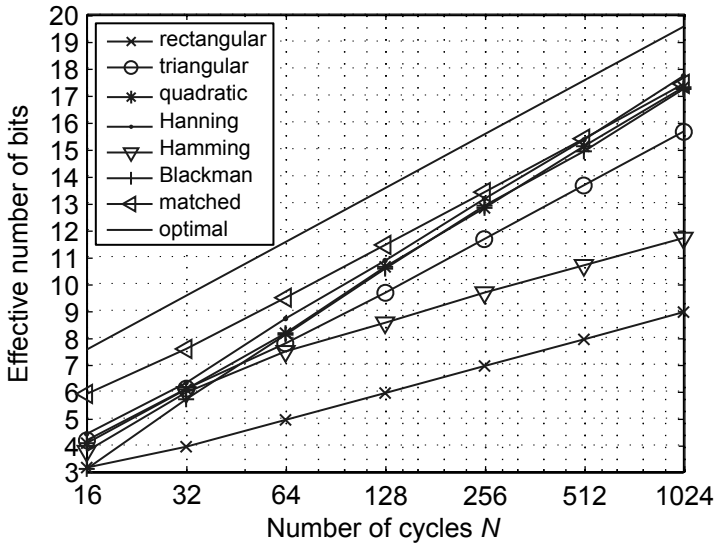


Figure 4.24. Effective number of bits obtained from a second-order $\Sigma\Delta$ converter as a function of the number of cycles N , for different decimation filters. The input range was restricted from 0.3 to 0.7, and b/a_1 was 2.

$\Sigma\Delta$ modulator: the quantization noise will be mainly located in the side lobes, which suggests that the filter's attenuation there should be maximized; however, if this results in a pass band that extends into the shaped quantization noise, the improvement may be undone. This suggests that there should be an optimal trade-off between pass-band width and side lobe attenuation. A further characteristic of interest is the roll-off towards $f_s/2$. Since the quantization noise of a $\Sigma\Delta$ modulator increases towards $f_s/2$, it can be expected that window functions with a steeper roll-off will perform better.

Figure 4.24 shows the ENOB that can be obtained from a second-order $\Sigma\Delta$ modulator with a decimation filter based on various commonly used window functions [22, 27]. The rectangular and Hamming windows have a 20 dB/dec roll-off. Since this is less than the 40 dB/dec increase of quantization noise (see Figure 4.20), they result in a less than 2 bits/dec increase of ENOB and are therefore not very interesting. The ENOB obtained with windows with a 60 dB/dec roll-off ('quadratic', Hanning, Blackman) increases by more than 2 bits/dec for small N , but eventually runs parallel to the ENOB of an optimal decimation filter, which, as was described in the previous section, increases by 2 bits/dec. The triangular window, which has a 40 dB/dec roll-off, maintains a 2 bits/dec increase for all N , and has a roughly 4 bits lower ENOB than the optimal filter. For comparison, the ENOB obtained with a matched decimation

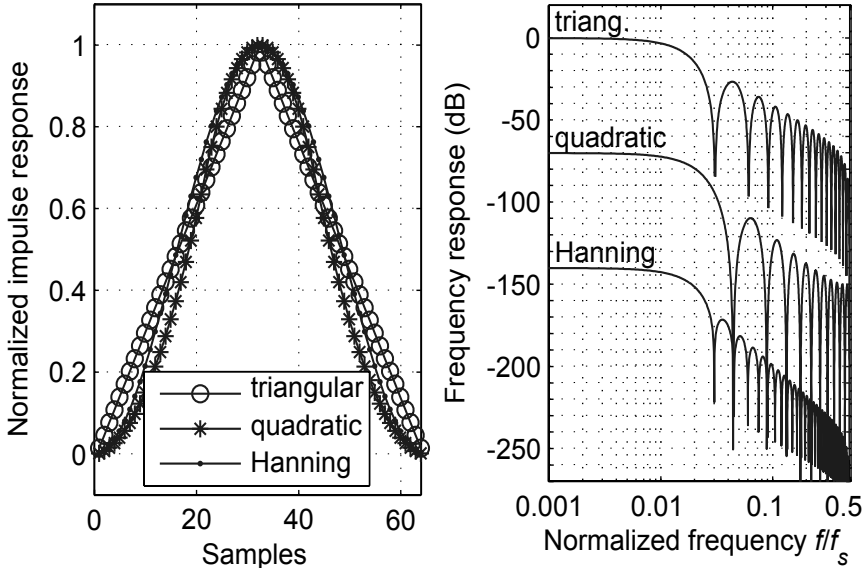


Figure 4.25. Impulse response (left) and frequency response (right) of a triangular window, a piecewise quadratic window, and a Hanning window; the frequency responses are offset for clarity.

filter is also shown. At higher values of N , it is comparable to that obtained with filters based on a window with 60 dB/dec roll-off.

The choice of window function is a trade-off between ENOB performance and implementation complexity. In terms of complexity, three window functions stand out: the rectangular, triangular and quadratic windows. The rectangular window of length N corresponds to a sinc filter. The triangular window can be seen as a convolution of two rectangular windows of length $N/2$, and therefore corresponds to a sinc^2 filter. A convolution of three rectangular windows of the length $N/3$ leads to a window with a piecewise quadratic shape (referred to a ‘quadratic’ in Figure 4.20), which corresponds to a sinc^3 filter. The impulse responses of these filters can be generated relatively easily using counters and accumulators, in contrast with the other windows, which require more complex hardware. Figure 4.25 shows, for illustration, the impulse response and frequency response of a triangular window, a quadratic window and a Hanning window.

Since a piecewise quadratic window results in an ENOB comparable to other more complex windows, it is an attractive choice. To obtain the ENOB of 15 bits required in a precision temperature sensor, about 512 clock cycles are needed.

The penalty of using an even simpler triangular window is an increase in the number of cycles to about 850.

4.5.3 Linear Scaling of the Conversion Result

In a temperature sensor, the conversion result D_{out} produced by the ADC should be readily interpretable as a temperature in degrees Celsius. Often, a binary fixed-point representation is used (see for instance [2]). Assuming the bit density μ of the bitstream is a linear function of temperature³, it has to be scaled linearly to arrive at D_{out} :

$$D_{out} = A \cdot \mu + B \quad (4.32)$$

If, for example, $\mu = T/600$ K, and an output format is desired in which the least significant bit corresponds to $\frac{1}{256}$ °C, the coefficients should be $A = 600 \cdot 256$ and $B = -273.15 \cdot 256$. While this scaling could be implemented by a digital multiplier that processes the output of the decimation filter, it would be advantageous in terms of chip area if the decimation filter itself could be designed to perform the scaling, and directly provide an output in the right format.

In theory, it is fairly straightforward to make a FIR decimation filter with a gain A and an offset B . Since such a decimation filter computes a weighted sum of the bitstream, where the weighting factors are given by the filter's impulse response, the gain factor A can be set by appropriate scaling the impulse response. For a given impulse response, the filter's gain can be found by evaluating its response to a bitstream consisting of only ones, and equals the sum of the impulse response values $h(n)$:

$$A = \sum_{i=1}^N h(i),$$

where N is the number of cycles. An offset B can easily be added to the filter's output.

In a practical implementation, the impulse response values have a finite resolution. As a result of rounding errors, the filter's gain will differ from the desired value, and, moreover, the filter's frequency response will deviate from the ideal response. Both will result in an increase in quantization errors of the overall converter. The magnitude of these errors can be evaluated by simulation. If necessary, the resolution of the impulse response values have to be increased.

³Some curvature correction techniques, such as the ratiometric correction, which uses a temperature-dependent reference, result in a bit density that is not exactly proportional to absolute temperature, but is still a linear function of temperature.

A way of fine-tuning the gain of a decimation filter is to add a lower-order filter in parallel to the main filter and add the outputs of the two filters. Consider, for example, a triangular decimation filter. If integer values are used for its impulse response, its gain is $N(N-1)/2 \simeq N^2/2$. This gain cannot be set very accurately by varying N . If a secondary filter based on a rectangular window with length $N_2 \leq N$ is used in parallel to this filter, the overall gain is $N^2/2 + N_2$, which can be accurately defined by choosing an appropriate value for N_2 .

Although the quantization errors at the output of the secondary filter are much larger than those at the output of the main filter, they need not significantly reduce the overall ENOB. This is because the secondary filter only determines a fraction $2N_2/N^2$ of the overall conversion result. Figure 4.24 shows that the ENOB obtained with a rectangular filter of length N_2 is roughly $\log_2(N_2) - 1$, which means that it has a peak quantization error of V_{REF}/N_2 . The relative error at the output due to the rectangular filter is then $2V_{REF}/N^2$. For $N = 850$, for instance, this is $2.77 \cdot 10^{-6} \cdot V_{REF}$. A triangular filter of length $N = 850$ leads to an ENOB of 15 bits, which corresponds to an error of $15.26 \cdot 10^{-6} \cdot V_{REF}$. The overall error is then $18.0 \cdot 10^{-6} \cdot V_{REF}$, which corresponds to an overall ENOB of 14.8. In other words, the ENOB does not decrease significantly as a result of the fine-tuning. This is confirmed by the simulation result shown in Figure 4.26.

4.5.4 Compensating for Non-Linearity

In Section 3.5, various curvature-correction techniques have been discussed that aim at linearizing the temperature dependency of the bit density μ . Any residual systematic non-linearity in this temperature dependency can, in principle, be removed in the digital domain. The output of the decimation filter could be fed into a lookup table that performs the required non-linear mapping, but this would require a very large ROM. A digital piece-wise linear or polynomial correction would be a more efficient implementation [28], but this would still require significant hardware overhead.

Malcovati *et al.* have presented a technique that can be used to efficiently implement a non-linear correction *inside* the decimation filter. Their filter was intended for use with a gas flow sensor with a very non-linear characteristic [29]. Their decimation filter is preceded by a sinc filter and a lookup table (Figure 4.27). The sinc filter produces an n_1 -bit filtered version of the bitstream. This filtered version is essentially a coarse, oversampled representation of the bit density μ . It is linearized by a lookup table with 2^{n_1} entries of n_2 bits each. The output of this lookup table, finally, is fed into the actual decimation filter, which decimates the linearized data and produces full-resolution output data.

Because the n_1 -bit wide intermediate data has a much smaller resolution than the final conversion result, the lookup table only requires a small ROM.

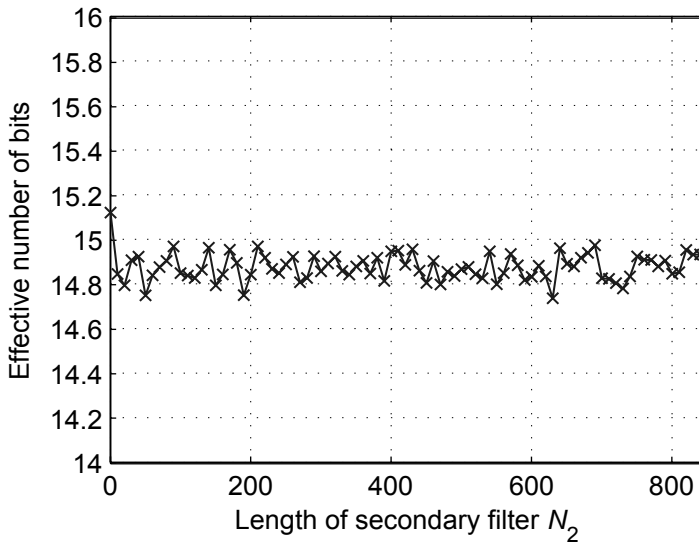


Figure 4.26. Simulated ENOB of a second-order modulator ($b/a_1 = 2$) with a triangular decimation filter of length $N = 850$ of which the gain is fine-tuned by means of a parallel rectangular filter of length N_2 .

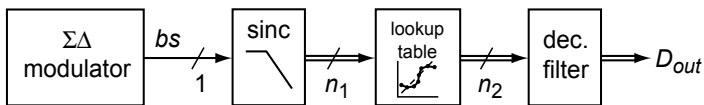


Figure 4.27. Non-linear decimation filter based on a lookup table that processes low-resolution oversampled data [29].

Nevertheless, a smooth linearization is obtained, because the lookup table acts on oversampled data. If, for example, 4-bit intermediate data is used, and the bit density is 0.3, the output of the sinc filter will switch between $4/16$ and $5/16$ in such a way that its average is 0.3. As a result, entries 4 and 5 of the lookup table will be addressed alternately, so that the average of its output (and hence the output of the decimation filter) is an interpolation between these two entries. The number of bits n_1 and n_2 are chosen such that the linearization is performed with sufficient accuracy.

In the application of Malcovati *et al.*, a fairly large non-linearity ($> 10\%$) had to be corrected. In a temperature sensor, in contrast, the non-linearity is typically much smaller. Depending on whether any curvature correction techniques are applied in the front-end circuitry, the non-linearity is between 0.1°C (0.016%)

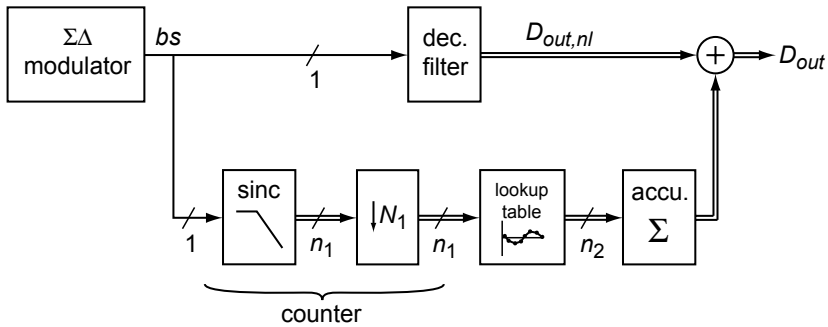


Figure 4.28. Decimation filter with an additive non-linear correction.

and 2°C (0.33%). In this case, it is more efficient to use an additive non-linear correction term, rather than the described multiplicative correction [29].

An implementation of such an additive correction using an oversampled lookup table is shown in Figure 4.28. The main decimation filter is, for instance, a sinc^2 filter, which produces a slightly non-linear temperature reading $D_{out,nl}$. The bitstream is also fed into a sinc filter of length N_1 followed by a decimator, which produces output data at a rate of f_s/N_1 . The output of the sinc filter is a coarse representation of temperature. A lookup table maps this temperature information onto a correction value. Throughout a conversion, the correction values are accumulated and finally added to the output of the main decimation filter. The result is a linearized output D_{out} .

The implementation of such an additive correction requires even less hardware than the multiplicative correction. The main decimation filter is much simpler, as it processes the bitstream, rather than the n_2 -bit wide output of the lookup table. Moreover, the sinc filter followed by a decimator can be implemented by means of a counter that counts the number of ones in every segment of N_1 bits of the bitstream. Finally, the entries in the lookup table now only contain the non-linear error, rather than the full transfer function.

Figure 4.29 shows the simulated non-linearity introduced using the described technique in a second-order $\Sigma\Delta$ modulator with a sinc^2 decimation filter with a length $N = 1024$. A lookup table with 32 entries ($n_1 = 5$) was used. The (arbitrary) 4-bit values ($n_2 = 4$) of these entries are shown in Figure 4.29a. The lookup table was preceded by a sinc filter with a length of 32, the output of which was decimated by a factor $N_1 = 32$. The non-linearity of the conversion result D_{out} as a function of the input level V_{IN}/V_{REF} is shown in Figure 4.29b. Note that the noise at the extremes of the input range is a result of instability of the modulator, and is unrelated to the correction technique. The result is clearly a smoothly interpolated version of the values stored in the lookup table.

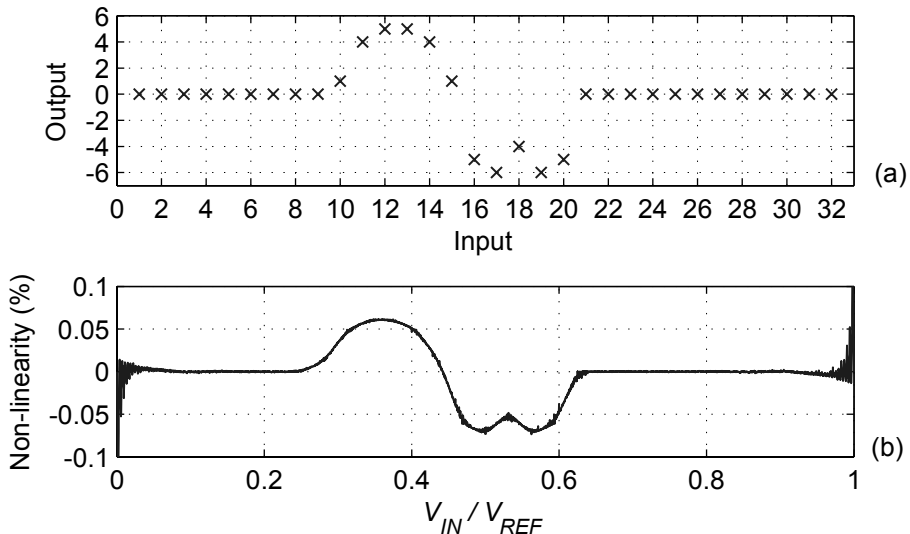


Figure 4.29. Simulation of the additive non-linear correction: (a) lookup table entries with an arbitrary correction function; (b) non-linearity of the output of a second-order $\Sigma\Delta$ modulator combined with the described non-linear decimation filter.

4.6 Filtering of Dynamic Error Signals

If dynamic error-correction techniques, such as chopping and dynamic element matching, are used in the front-end circuitry of a $\Sigma\Delta$ converter, dynamic error signals are present in the input voltage V_{IN} and the reference voltage V_{REF} of the $\Sigma\Delta$ modulator. This section discusses how to organize the timing of these error signals so as to ensure that these error signals are filtered out and an accurate conversion result is obtained.

4.6.1 Normal-Mode Rejection

The low-pass characteristic of a $\Sigma\Delta$ converter can be used to suppress disturbances present in its input signal. Such disturbances include, for instance, cross-talk from the mains, but also dynamic error signals generated in the front-end circuitry. The latter is illustrated in Figure 4.30, which shows a $\Sigma\Delta$ modulator with front-end circuitry. The front-end circuitry employs dynamic error correction (DEC) techniques to provide an input voltage V_{IN} and reference voltage V_{REF} of which the average value is accurate. The dynamic error signals superimposed on this average value have to be filtered out by the $\Sigma\Delta$ modulator and/or the decimation filter.

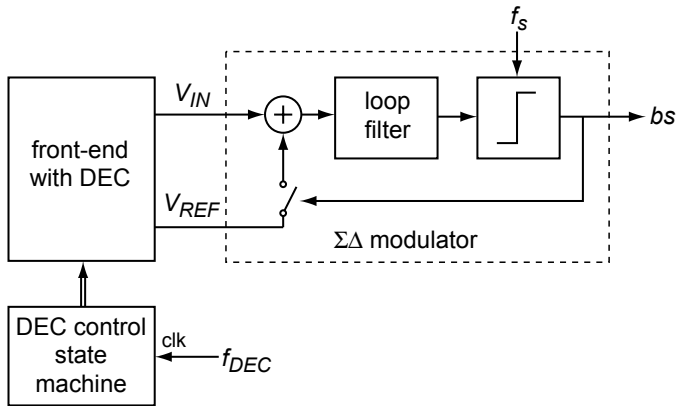


Figure 4.30. Sigma-delta modulator with front-end circuitry that generates the modulator's input and reference voltage using dynamic error correction (DEC) techniques.

The frequency content of the dynamic error signals depends on the timing of the switching in the front-end circuitry. In Figure 4.30, this timing is controlled by a state machine. If, for example, dynamic element matching is applied, this state machine could be a counter that cyclically selects the element that have to be matched. The frequency f_{DEC} at which the state machine operates thus determines the frequency of the dynamic error signals.

Dynamic error signals can be suppressed by ensuring that their frequency corresponds to a zero in the frequency response of the converter. These zeros are mainly determined by the frequency response of the decimation filter. The zeros of a triangular filter with an impulse response of length N occur at

$$f = k \cdot \frac{2f_s}{N} \quad k = 1, 2, 3, \dots \quad (4.33)$$

where f_s is the sampling frequency (see Figure 4.23). Note that there is no zero at f_s/N (as would be the case with a rectangular filter of the same length). It is generally attractive to choose a low frequency for dynamic error signals, as this minimizes the errors and power consumption associated with switching in the front-end [30].

The simulated response of a complete second-order $\Sigma\Delta$ converter to periodic errors in V_{IN} and in V_{REF} is shown in Figure 4.31. The simulated converter consists of a second-order modulator and a triangular decimation filter with $N = 512$. The error signals used were square waves with an amplitude of 10% of V_{IN} or V_{REF} . Their frequency was swept from $f_s/(2N)$ to $f_s/2$ in binary steps. Error signals in V_{IN} are suppressed to the level of the converter's quantization error, as expected based on (4.33). Error signals in V_{REF} , in contrast, are only suppressed poorly.

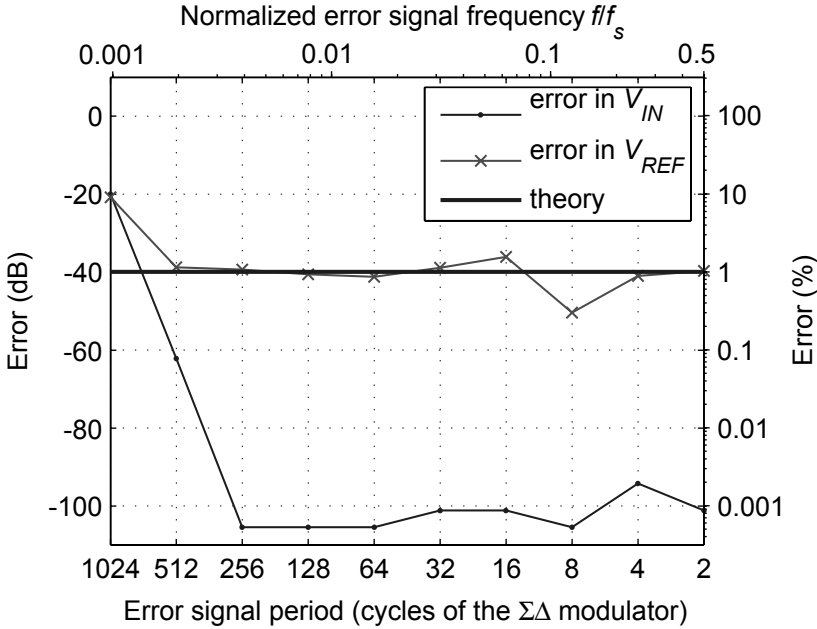


Figure 4.31. Simulated error in the output of a second-order $\Sigma\Delta$ converter with a triangular decimation filter when a square wave error signal with an amplitude of 10% is present in either V_{IN} or V_{REF} . The bold line is the limit predicted by (4.34). Simulation parameters: $b/a_1 = 2$, $V_{IN}/V_{REF} = 0.54321$ and $N = 512$.

The poor suppression of error signals in V_{REF} can be explained as follows [10]. Suppose that the reference voltage switches back and forth between $V_{REF} + V_{error}$ and $V_{REF} - V_{error}$ at a frequency of $2f_s/N$, which corresponds to the first zero of the triangular decimation filter. The modulator's bit density, which should ideally be V_{IN}/V_{REF} , now switches between $V_{IN}/(V_{REF} + V_{error})$ and $V_{IN}/(V_{REF} - V_{error})$. The average bit density is therefore

$$\begin{aligned} \mu &= \frac{1}{2} \left(\frac{V_{IN}}{V_{REF} + V_{error}} + \frac{V_{IN}}{V_{REF} - V_{error}} \right) \\ &= \frac{V_{IN}}{V_{REF} - V_{error}^2/V_{REF}}. \end{aligned} \quad (4.34)$$

This shows that a residual error of V_{error}^2/V_{REF} remains. This level is indicated by the bold line in Figure 4.31, and is in agreement with the simulation results.

This residual error cannot be eliminated by using a frequency close to $f_s/2$, because the dynamic error signals then coincide with the shaped quantization noise of the modulator. Since dynamic error signals in V_{REF} are *multiplied* by the bitstream, quantization noise may be down-converted to DC [31]. Thus, the

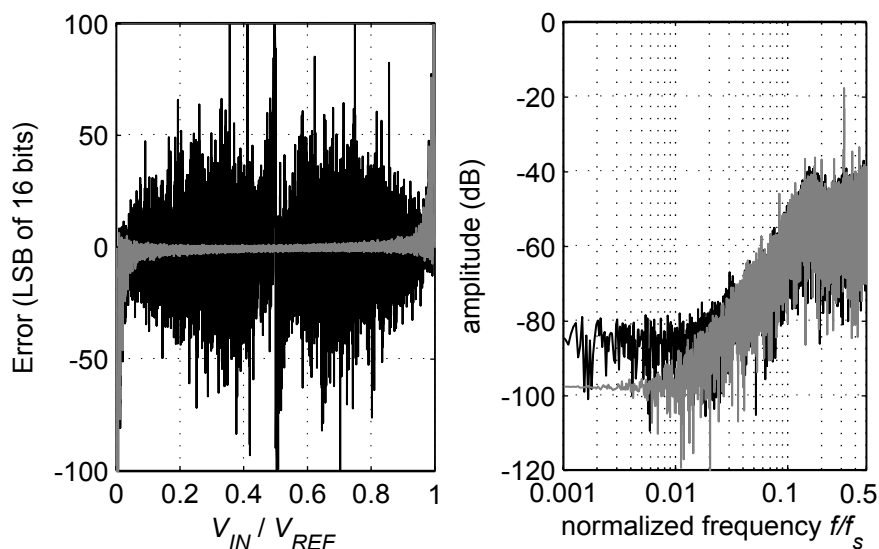


Figure 4.32. Quantization error (left) and bitstream spectrum (right) of a second-order modulator, with a $\pm 1\%$ error signal in V_{REF} at $f_s/2$ (black), and with a clean V_{REF} (gray); the quantization error was determined using a triangular filter ($N = 1024$).

dynamic error correction is rendered ineffective, and the in-band quantization noise is increased. This is illustrated in Figure 4.32, which shows a strong increase in quantization error, and an associated increase in low-frequency noise in the bitstream, if a $\pm 1\%$ error signal at a frequency of $f_s/2$ is present in V_{REF} .

Residual errors and increased noise resulting from error signals in V_{REF} can be prevented by modulating these error signals to a multiple of f_s [32]. They will then be averaged out in the loop filter within every cycle of the $\Sigma\Delta$ modulator. A disadvantage of this approach is that the loop filter has to be able to handle these faster signals. In a switched-capacitor implementation, this means that the first integrator of the loop filter will have to sample the input and reference voltages more than once per $\Sigma\Delta$ cycle, which will result in an increase in power consumption. More efficient alternatives will be discussed in the next section.

4.6.2 Bitstream-Controlled Timing of Dynamic Error Signals

The main problem with dynamic error signals near $f_s/2$ in the reference of a $\Sigma\Delta$ modulator, is that they may be correlated to the bitstream of the modulator, and will be modulated back to DC when multiplied by that bitstream. A possible solution is therefore to randomize the frequency of the dynamic error signals

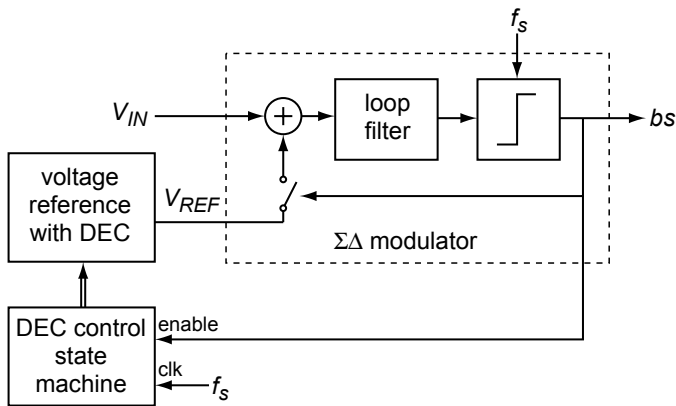


Figure 4.33. Bitstream-controlled timing of dynamic error correction in the voltage reference generator of a $\Sigma\Delta$ modulator.

[33]. This can be done, for instance, by using a pseudo-random clock to control the DEC state machine. Thus, intermodulation products will not end up at DC, but will be spread out around DC, resulting in a slight increase in noise, but not in large DC errors. In addition to the increased noise, a disadvantage of this technique is that a pseudo-random generator is needed.

An alternative way of preventing DC errors as a result of the multiplication of a dynamic error signal and the bitstream, is to derive the timing of the error signal from the bitstream [34, 35]. The principle of this ‘bitstream-controlled’ timing is illustrated in Figure 4.33. The DEC state machine is equipped with an enable input and is clocked at f_s . The enable input is driven by the bitstream bs , so that the state machine only proceeds to the next state if bs is one, which is when V_{REF} is actually applied to the loop filter. When bs is zero, the state machine is frozen. As result, the errors in the samples of V_{REF} during clock cycles in which bs is one will average out. The multiplication of V_{REF} by the bitstream does not disturb this property, and hence residual errors or increased quantization noise are prevented.

Bitstream-controlled timing is similar to the data-weighted averaging techniques used in multi-bit $\Sigma\Delta$ modulators, e.g. [36]. There, however, the goal is to linearize a multi-bit DAC, while here, dynamic error correction is applied to make the absolute value of a single-bit DAC accurate.

Figure 4.34 shows how bitstream-controlled timing works out for dynamic error correction with 2 states. In that case, the front-end circuitry generates a reference voltage that alternates between $V_{REF} + V_{error}$ and $V_{REF} - V_{error}$. The state machine that controls this switching only changes its state at the end of a clock cycle when bs is one. When bs is zero, the state machine is frozen

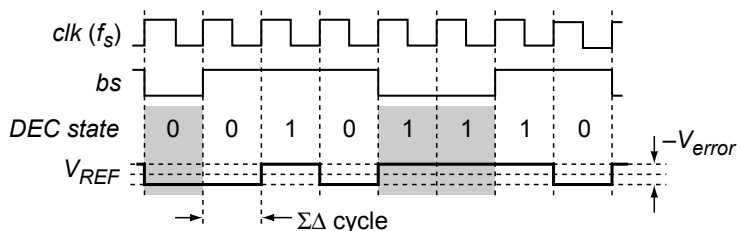


Figure 4.34. Bitstream-controlled timing in the case of two-phase dynamic error correction.

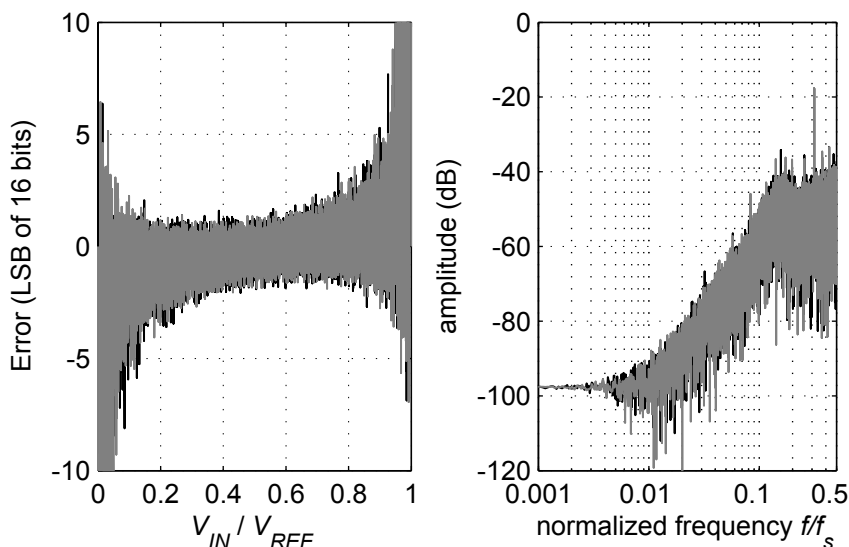


Figure 4.35. Quantization error (left) and bitstream spectrum (right) of a second-order modulator, with a $\pm 1\%$ error signal in V_{REF} with bitstream-controlled timing (black), and with a clean V_{REF} (gray); the quantization error was determined using a triangular filter ($N = 1024$).

(indicated by the gray boxes in Figure 4.34). Figure 4.35 shows the simulated quantization error as a function of the DC input level for an error signal of $\pm 1\%$. The quantization error is not increased compared to that of a converter with a clean reference. The same conclusion can be drawn from the spectrum of the bitstream.

In the case of the charge-balancing scheme of Figure 4.2b, there is no explicit reference voltage. V_{BE} and ΔV_{BE} are multiplied by the bitstream and its inverse, so that error signals in both these voltages can be modulated back to DC. Bitstream-controlled timing can be used to prevent this, as shown in Figure 4.36 [34]. There are now two DEC state machines, one which controls DEC

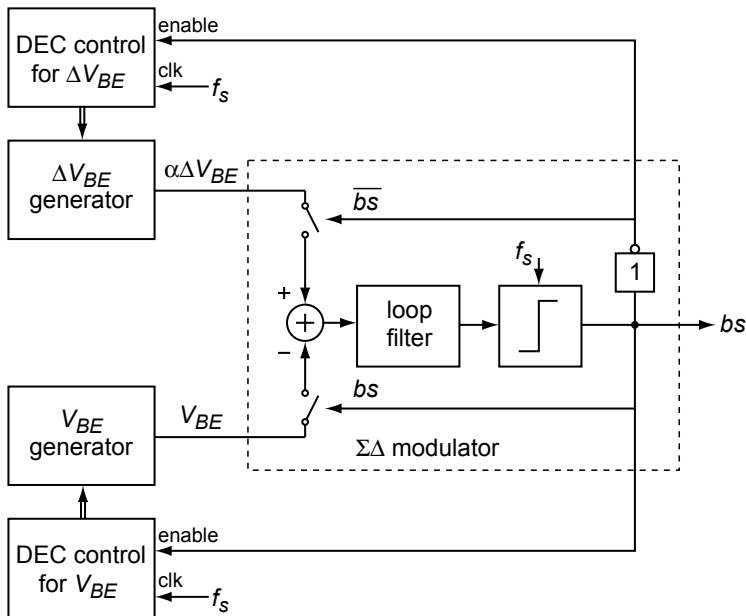


Figure 4.36. Bitstream-controlled timing of dynamic error correction applied to the temperature sensor front-end of Figure 4.2b.

in the circuitry that generates V_{BE} , and one that controls DEC in the circuitry that generates ΔV_{BE} . The former is enabled only when the bitstream is one, and the latter only when the bitstream is zero.

In the general case of a $\Sigma\Delta$ converter with a bipolar input range, the reference voltage is multiplied by $+1$ or -1 if the bitstream is 1 or 0, respectively. In that case, again two DEC state machines are needed, one of which is enabled by the bitstream and one by its inverse (Figure 4.37). The bitstream also determines which state machine provides, via a multiplexer, the control signals to the front-end circuitry. Thus, independent sequences of control signals are generated for the zeros and for the ones in the bitstream. As result, the errors in the samples of V_{REF} during clock cycles in which bs is one will average out, and similarly the errors during the clock cycles in which bs is zero will average out. The multiplication of V_{REF} by the bitstream does not disturb this property, and hence residual errors or increased quantization noise are prevented.

Bitstream-controlled timing does not have the disadvantages of the techniques discussed earlier. The frequency of the dynamic error signals applied to the modulator is never larger than $f_s/2$, so that the speed and hence the power consumption of the first integrator do not have to be increased. The residual error that occurs when switching at low frequencies does not occur. Finally,

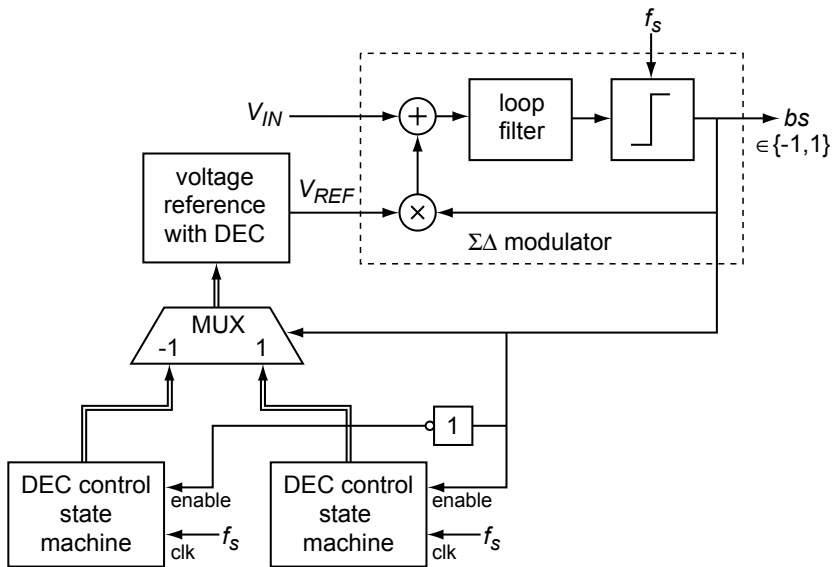


Figure 4.37. Bitstream-controlled timing applied to a $\Sigma\Delta$ modulator with a bipolar input range.

the increase in noise associated with pseudo-random timing is avoided, while the added circuitry only consists of some extra digital logic.

4.7 Conclusions

In this chapter, the design of $\Sigma\Delta$ ADCs for use in smart temperature sensors has been described. The ADC in such a sensor uses V_{BE} and ΔV_{BE} , which are generated using the techniques described in the previous chapter, to produce a temperature reading in a standardized digital format. In a precision sensor, the ADC should not introduce any significant errors in doing so. This means that it has to perform a 15-bit accurate conversion in about 100 ms.

A comparison of various classes of ADCs has shown that indirect ADCs are a good choice for this application. Such ADCs do not require precisely matched components, so that they can be implemented in CMOS without trimming. They are based on a modulator that converts the ratio of the input and reference voltage to an intermediate time-domain signal. It has been shown that the ratio required in a temperature sensor can be implemented using simple charge balancing between V_{BE} and ΔV_{BE} . The modulator of an indirect ADC can be either free-running, or synchronized to a clock. Synchronous modulators are preferred, because of their compatibility with the on-chip digital circuitry that converts the intermediate time-domain signal to a final temperature reading. Moreover, they are less sensitive to interference than free-running converters.

Within the class of synchronous indirect ADCs, $\Sigma\Delta$ ADCs are the most widely used. They consist of a $\Sigma\Delta$ modulator and a decimation filter. The modulator produces a bitstream proportional to the input signal, and the decimation filter converts this bitstream to a final reading. When using a higher-order $\Sigma\Delta$ modulator, relatively high resolutions can be obtained in a limited number of clock cycles. While a first-order $\Sigma\Delta$ modulator needs more than 32000 clock cycles to obtain a resolution of 15 bits, the same resolution can be obtained with a second-order modulator using less than 1000 cycles. Thus, modest clock frequencies in the order of 10 kHz can be used, which is advantageous in terms of power consumption.

The characteristics and limitations of both first-order and second-order $\Sigma\Delta$ modulators have been discussed. The resolution obtainable with these modulators is not only determined by the number of cycles, but also by the DC gain of the integrators used. For an n -bit resolution, the DC gain of the integrator in a first-order modulator has to be in the order of 2^n . Roughly the same holds for the combined DC gain of the integrators in a second-order modulator.

The performance of a second-order modulator is also affected by the topology and the coefficients of its loop filter. A single-loop topology with feedback to the second integrator has been shown to be an attractive choice in view of its compatibility with the simple charge balancing scheme mentioned above. A disadvantage of a single-loop topology is that the integrator outputs will clip if the input signal approaches the extremes of the input range. The mentioned charge-balancing scheme, however, conveniently avoids these parts of the inputs range. For the middle of the input range, which is the part used, the feedback coefficient can be used to trade resolution against integrator output swing.

A $\Sigma\Delta$ ADC in a temperature sensor is operated as an incremental converter, which means that it is reset at the beginning of a conversion, runs for N cycles to produce a conversion result, and then typically powers down to save power. As a result, the decimation filter necessarily has a finite impulse response of length N . Such filters can be designed based on window functions. For a first-order modulator, a rectangular window results in an optimal resolution. For a second-order modulator, windows of which the frequency response rolls off by 60 dB/dec perform best. A piecewise quadratic window, which has a sinc^3 response, is an example of such a window. It achieves a 15-bit resolution in about 512 cycles. To simplify the implementation, a triangular window (with a sinc^2 response) can be used at the expense of some loss of resolution. The number of cycles then has to be increased to about 850 to achieve 15 bits.

It has been shown how a parallel combination of decimation filters can be used to accurately define the gain of the ADC, so as to directly produce a conversion result scaled to degrees Celsius. The gain of a triangular filter, for example, can be fine-tuned by adding the result of rectangular filter to its output. It has also been shown how an oversampled lookup table can be used

to generate a programmable non-linearity in the decimation with minimum hardware overhead. This can be used to provide curvature correction in a temperature sensor.

Finally, it has been discussed how the ADC can be designed to filter dynamic error signals from its input. This is important in a precision temperature sensor, since the front-end circuitry typically relies heavily on dynamic error correction techniques to generate accurate voltages V_{BE} and ΔV_{BE} . This is the reason why decimation filters matched to the loop filter, which are typically used in incremental $\Sigma\Delta$ ADCs to obtain a relatively high resolution, are not an attractive choice in a smart temperature sensor: they result in a poor rejection of error signals. Window-based decimation filters, in contrast, can have notches in their frequency response. If the frequencies of the dynamic error signals correspond to such notches, they are effectively filtered out.

Special care has to be taken if dynamic error signals get multiplied by the bitstream of the modulator. This happens if they are present in the reference of the modulator. It has been shown that such signals are not filtered out completely if they have a low frequency compared to the sampling frequency of the modulator. If their frequency is close to half the sampling frequency, they can result in down-conversion of quantization noise. These problems can be solved by letting the timing of the dynamic error correction depend on the bitstream. Such bitstream-controlled timing requires little extra circuitry, and ensures that error signals in the reference are properly filtered out.

This chapter has dealt with system-level issues in the design of $\Sigma\Delta$ ADCs for temperature sensors. The biggest challenge in the circuit implementation of such ADCs is the design of the first integrator of the loop filter. It was already mentioned in this chapter that there are two flavours of loop filters: continuous-time and switched-capacitor filters. The corresponding two integrator types will be discussed in detail in the next chapter.

References

- [1] *IEEE standard for terminology and test methods for analog-to-digital converters*, IEEE Std. 1241-2000, Dec. 2000.
- [2] “LM75 data sheet,” National Semiconductor Corp., Feb. 2004, www.national.com.
- [3] P. R. van der Meer, G. C. M. Meijer, M. J. Vellekoop, H. M. M. Kerkvliet, and T. J. J. van den Boom, “A temperature-controlled smart surface-acoustic-wave gas sensor,” *Sensors and Actuators*, vol. 71, no. 1-2, pp. 27–34, Nov. 1998.
- [4] R. J. van de Plassche, *CMOS Integrated Analog-to-Digital and Digital-to-Analog Converters*, 2nd ed. Boston: Kluwer Academic Publishers, 2003.
- [5] G. C. M. Meijer *et al.*, “A three-terminal integrated temperature transducer with micro-computer interfacing,” *Sensors and Actuators*, vol. 18, pp. 195–206, June 1989.

- [6] P. Krummenacher and H. Oguey, "Smart temperature sensor in CMOS technology," *Sensors and Actuators*, vol. A21-A23, pp. 636–638, Mar. 1990.
- [7] F. R. Riedijk and J. H. Huijsing, "An integrated absolute temperature sensor with sigma-delta A-D conversion," *Sensors and Actuators*, vol. 34, pp. 249–256, Sept. 1992.
- [8] A. Bakker and J. H. Huijsing, "Micropower CMOS temperature sensor with digital output," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 7, pp. 933–937, July 1996.
- [9] M. Tuthill, "A switched-current, switched-capacitor temperature sensor in 0.6- μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 7, pp. 1117–1122, 1998.
- [10] A. Bakker and J. H. Huijsing, *High-Accuracy CMOS Smart Temperature Sensors*. Boston: Kluwer Academic Publishers, 2000.
- [11] F. M. L. van der Goes and G. C. M. Meijer, "Sigma-delta versus oscillator-based converters in low-cost accurate sensor systems," in *Proc. IMTC*, June 1996, pp. 1151–1153.
- [12] S. Ouzounov, E. Roza, H. Hegt, G. van der Weide, and A. van Roermund, "Design of high-performance asynchronous sigma delta modulators with a binary quantizer with hysteresis," in *Proc. CICC*, Oct. 2004, pp. 181–184.
- [13] S. R. Norsworthy, R. Schreier, and G. C. Temes, Eds., *Delta-Sigma Data Converters: Theory, Design and Simulation*. Piscataway, New York: IEEE Press, 1997.
- [14] O. Bajdechi and J. H. Huijsing, *Systematic Design of Sigma-Delta Analog-to-Digital Converters*. Boston: Kluwer Academic Publishers, 2004.
- [15] R. Schreier and B. Zhang, "Delta-sigma modulators employing continuous-time circuitry," *IEEE Transactions on Circuits and Systems—Part I: Fundamental Theory and Applications*, vol. 43, no. 4, pp. 324–332, Apr. 1996.
- [16] R. J. van der Plassche, "A sigma-delta modulator as an A/D converter," *IEEE Transactions on Circuits and Systems*, vol. 25, no. 7, pp. 510–514, July 1978.
- [17] J. Márkus, J. Silva, and G. C. Temes, "Theory and applications of incremental $\Sigma\Delta$ converters," *IEEE Transactions on Circuits and Systems—Part I: Fundamental Theory and Applications*, vol. 51, no. 4, pp. 678–690, Apr. 2004.
- [18] J. Robert, G. C. Temes, V. Valencic, R. Dessoulavy, and P. Deval, "A 16-bit low-voltage CMOS A/D converter," *IEEE Journal of Solid-State Circuits*, vol. SC-22, no. 2, pp. 157–163, Apr. 1987.
- [19] J. Robert and P. Deval, "A second-order high-resolution incremental A/D converter with offset and charge injection compensation," *IEEE Journal of Solid-State Circuits*, vol. 23, no. 3, pp. 736–741, June 1988.
- [20] O. Feely and L. O. Chua, "The effect of integrator leak in $\Sigma\text{-}\Delta$ modulation," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 11, pp. 1293–1305, Nov. 1991.
- [21] P. C. de Jong, G. C. M. Meijer, and A. H. M. van Roermund, "A new dithering method for sigma-delta modulators," *Analog Integrated Circuits and Signal Processing*, vol. 10, no. 3, pp. 193–204, Aug. 1996.

- [22] J. C. Candy, "A use of double integration in sigma delta modulation," *IEEE Transactions on Communications*, vol. COM-33, no. 3, pp. 249–258, Mar. 1985.
- [23] B. E. Boser and B. A. Wooley, "The design of sigma-delta modulation analog-to-digital converters," *IEEE Journal of Solid-State Circuits*, vol. SC-23, no. 6, pp. 1298–1308, Dec. 1988.
- [24] G. Temes and J. Steensgaard, "Structural optimization and scaling of delta-sigma modulators," in *Lecture Notes of the EPFL Advanced Engineering Course on Delta Sigma Converters for Telecom*, 2000.
- [25] C. Lyden, U.S. Patent 5 189 419, Feb. 23, 1993.
- [26] C. Lyden, J. Ryan, C. A. Ugarte, J. Kornblum, and F. M. Yung, "A single shot sigma delta analog to digital converter for multiplexed applications," in *Proc. CICC*, May 1995, pp. 203–206.
- [27] A. W. M. van den Enden and N. A. M. Verhoeckx, *Discrete-time signal processing*. Upper Saddle River, NJ, USA: Prentice Hall, 1989.
- [28] G. v. d. Horn and J. H. Huijsing, *Integrated Smart Sensors: Design and Calibration*. Boston: Kluwer Academic Publishers, 1998.
- [29] P. Malcovati, C. A. Leme, P. O'Leary, F. Maloberti, and H. Baltes, "Smart sensor interface with A/D conversion and programmable calibration," *IEEE Journal of Solid-State Circuits*, vol. 29, no. 8, pp. 963–966, Aug. 1994.
- [30] W. Lee, "A 4-channel, 18b $\Sigma\Delta$ modulator IC with chopped-offset stabilization," in *Dig. Techn. Papers ISSCC*, Feb. 1996, pp. 238–239.
- [31] Y.-C. Huang and W.-S. Wey, "Second-order delta-sigma modulation with interfered reference," *IEEE Transactions on Circuits and Systems—Part II: Analog and Digital Signal Processing*, vol. 48, no. 2, pp. 192–197, Feb. 2001.
- [32] B. P. D. Signore, D. A. Kerth, N. S. Sooch, and E. J. Swanson, "A monolithic 20-b delta-sigma A/D converter," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 6, pp. 1311–1317, Dec. 1990.
- [33] C. B. Wang, "A 20-bit 25-khz delta-sigma A/D converter utilizing a frequency-shaped chopper stabilization scheme," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 3, pp. 566–569, Mar. 2001.
- [34] M. A. P. Pertijs and J. H. Huijsing, "A sigma-delta modulator with bitstream-controlled dynamic element matching," in *Proc. ESSCIRC*, Sept. 2004, pp. 187–190.
- [35] M. A. P. Pertijs, K. A. A. Makinwa, and J. H. Huijsing, "Bitstream controlled reference signal generation for a sigma-delta modulator," U.K. Patent Application 0 411 884.0, 2004.
- [36] B. C. Leung and S. Sutarja, "Multibit $\Sigma - \Delta$ A/D converter incorporating a novel class of dynamic element matching techniques," *IEEE Transactions on Circuits and Systems—Part II: Analog and Digital Signal Processing*, vol. 39, no. 1, pp. 35–51, Jan. 1992.

Chapter 5

PRECISION CIRCUIT TECHNIQUES

This chapter discusses the circuit implementation of CMOS smart temperature sensors. More specifically, it focusses on the implementation of accurate charge balancing in the sigma-delta ADC. Implementations based on continuous-time and switched-capacitor circuits are discussed. Their performance in terms of noise, accuracy and power consumption is analyzed, and solutions to mismatch- and offset-related errors are presented.

5.1 Introduction

In the previous chapter, it was shown how a $\Sigma\Delta$ ADC can be used in a smart temperature sensor to convert the voltages V_{BE} and ΔV_{BE} into a digital temperature reading. The biggest challenge in the circuit implementation of such an ADC lies in the design of its first integrator. In this integrator, V_{BE} and ΔV_{BE} are integrated in such a way that the charge contributed by one of them is balanced out by the other (see Figure 4.2b). While errors introduced in later stages of the ADC are attenuated by the gain of the first integrator, errors introduced in the first integrator directly add to V_{BE} and ΔV_{BE} , and thus degrade the sensor's accuracy [1].

5.1.1 Methods of Voltage-to-Charge Conversion

In order to integrate the voltages V_{BE} and ΔV_{BE} , they have to be converted into charge. There are essentially two ways of doing this: by sampling them on a capacitor, or by converting them into a current using a resistor [1, 2].

The first approach leads to a switched-capacitor (SC) integrator, the principle of which is shown in Figure 5.1a. In a first phase ϕ_1 , the input voltage (V_{BE} or ΔV_{BE}) is sampled on a capacitor C_S and is thus converted to a charge $C_S \cdot V_{IN}$. In a second phase ϕ_2 , this charge is added to an integration capacitor C_{int} ,

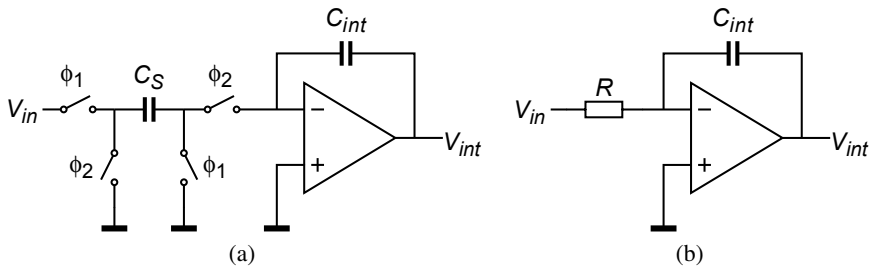


Figure 5.1. (a) Basic switched-capacitor integrator and (b) basic continuous-time integrator.

leading to an integrator coefficient that is defined by the ratio of the sampling and the integration capacitor:

$$a_{SC} = \frac{C_S}{C_{int}}. \quad (5.1)$$

The second approach leads to a continuous-time (CT) integrator, which is shown in Figure 5.1b. The input voltage is applied across a resistor R . As a result, a current V_{in}/R flows into the integration capacitor C_{int} during a time interval Δt , leading to an integrator coefficient of

$$a_{CT} = \frac{\Delta t}{RC_{int}}. \quad (5.2)$$

In sections 5.2 and 5.3, these two approaches will be analyzed in terms of their accuracy, noise and power consumption. First, however, an upper bound on the power consumption of the sensor will be established based on errors due to self-heating. Then, to aid the analysis in the following sections, a method of calculating the output-referred noise based on analysis of a single integration cycle of a $\Sigma\Delta$ modulator will be introduced. Finally, to establish a lower bound on the noise performance, the noise contribution of the bipolar front-end will be analyzed.

5.1.2 Maximum Power Consumption Based on Self-Heating

As will become clear in the following sections, the noise and accuracy performance of a given readout circuit can often be improved at the expense of higher bias currents, and, in consequence, a higher power consumption. Power consumption, however, cannot be increased indefinitely, because it leads to self-heating, which in turn results in a measurement error. How much a sensor's temperature will rise as a result of a given amount of power dissipation depends on many factors, such as the die size, the type of package, and the way the package is mounted [3,4].

Table 5.1. Typical values for the thermal resistance for several packages, along with the associated maximum supply currents; data obtained from [6] and [7].

package	thermal res. die to ambient θ_{JA} (K / W)	supply current that causes 0.01 K of self heating in still air (μA)		
		$V_{DD} = 2.5$ V	3.3 V	5.5 V
plastic DIP8	110	36	27	16
ceramic DIP8	110	36	27	16
plastic SO8	220	18	13	8
plastic TO-92	180	22	16	10
plastic SOT-23	180	22	16	10
metal can TO-46	400	10	7	4

An estimate of the worst-case self-heating can be obtained from the thermal resistance θ_{JA} from the die to its environment¹. This is typically measured by mounting a chip on a test board and measuring its temperature rise in still air [5]. Typical values of θ_{JA} are given in Table 5.1 for various packages commonly used for integrated temperature sensors. If the dissipated power is P , the steady-state temperature rise ΔT will be

$$\Delta T = P \cdot \theta_{JA}. \quad (5.3)$$

Using $P = V_{DD} \cdot I_{supply}$, the supply current I_{supply} can be calculated that causes a temperature rise ΔT for a given supply voltage V_{DD} and thermal resistance θ_{JA} :

$$I_{supply} = \frac{\Delta T}{\theta_{JA} \cdot V_{DD}}. \quad (5.4)$$

Table 5.1 lists the supply currents that cause 0.01 K of self-heating for various common supply voltages. For most packages, the supply current has to be limited to a few tens of μA in order to keep the self-heating in the order of 0.01 K.

Self-heating can be reduced by powering down a sensor in between measurements, rather than operating it continuously. In that case, the supply current during operation can be higher than the values mentioned in Table 5.1. For example, if a sensor is powered down 90% of the time, its average power consumption is reduced by a factor of ten². The temperature error due to self-heating is then at best also reduced by a factor of ten. The exact reduction

¹The 'JA' in θ_{JA} stands for 'junction to ambient'.

²This is under the assumption that the sensor is completely powered down. In practice, often an oscillator and some digital logic remain active to decide when the next measurement has to be performed.

depends on the thermal capacitance of the sensor [4]. This capacitance forms a thermal low-pass filter that averages out the switched power dissipation of the chip. The associated ripple in the chip's temperature has to be taken into account when determining the residual temperature rise during a measurement.

5.1.3 Per-Cycle Analysis of Noise

In this section, the relation between the noise accumulated during a complete $\Sigma\Delta$ AD conversion and that accumulated during an individual $\Sigma\Delta$ cycle will be derived. Using this relation, the noise of CT and SC readout circuits can be predicted in the following sections based on an analysis of the noise of a single $\Sigma\Delta$ cycle.

During every cycle, the first integrator of a $\Sigma\Delta$ modulator accumulates either a charge $Q_{\Delta V_{BE}}$ proportional to ΔV_{BE} , or a charge $-Q_{V_{BE}}$ proportional to V_{BE} (see the charge-balancing scheme in Figure 4.2b). The total accumulated charge Q_{acc} after N clock cycles is then

$$Q_{acc} = N \{ (1 - \mu) \cdot Q_{\Delta V_{BE}} - \mu \cdot Q_{V_{BE}} \} + q_{n,acc}, \quad (5.5)$$

where μ is the average value of the bitstream (i.e. the fraction of cycles during which the bitstream was one), and $q_{n,acc}$ is the accumulated noise charge. The charges $Q_{\Delta V_{BE}}$ and $Q_{V_{BE}}$ in this equation are assumed to be noise-free. If the rms noise associated with $Q_{\Delta V_{BE}}$ and $Q_{V_{BE}}$ is $q_{n,\Delta V_{BE}}$ and $q_{n,V_{BE}}$, respectively, the accumulated noise can be expressed as:

$$q_{n,acc}^2 = N \{ (1 - \mu) \cdot q_{n,\Delta V_{BE}}^2 + \mu \cdot q_{n,V_{BE}}^2 \}. \quad (5.6)$$

The feedback in a $\Sigma\Delta$ converter acts so as to null the accumulated charge Q_{acc} . In practice, Q_{acc} will not be exactly zero at the end of a conversion. As discussed in the previous chapter, this leads to a quantization error. For now, this effect will be ignored, so that Q_{acc} can be assumed to be zero. Solving (5.5) for μ then gives:

$$\mu = \frac{Q_{\Delta V_{BE}}}{Q_{\Delta V_{BE}} + Q_{V_{BE}}} + \frac{q_{n,acc}}{(Q_{\Delta V_{BE}} + Q_{V_{BE}}) \cdot N}. \quad (5.7)$$

where the first term is the desired average value of μ , and the second term the output-referred accumulated noise. The standard deviation of the latter can be found by substituting equation (5.6):

$$\sigma_{\mu} = \frac{1}{Q_{\Delta V_{BE}} + Q_{V_{BE}}} \sqrt{\frac{(1 - \mu) \cdot q_{n,\Delta V_{BE}}^2 + \mu \cdot q_{n,V_{BE}}^2}{N}}. \quad (5.8)$$

Since ΔV_{BE} is usually an order of magnitude smaller than V_{BE} , noise introduced in the readout circuitry usually has the largest effect on ΔV_{BE} . Therefore,

equation (5.8) can be simplified by assuming that $q_{n,V_{BE}}$ is much smaller than $q_{n,\Delta V_{BE}}$:

$$\sigma_\mu \simeq \frac{q_{n,\Delta V_{BE}}}{Q_{\Delta V_{BE}} + Q_{V_{BE}}} \sqrt{\frac{(1-\mu)}{N}}, \quad (5.9)$$

which can be translated into degrees Celsius using equation (3.7):

$$\sigma_T = A \cdot \sigma_\mu = A \cdot \frac{q_{n,\Delta V_{BE}}}{Q_{\Delta V_{BE}} + Q_{V_{BE}}} \sqrt{\frac{(1-\mu)}{N}}. \quad (5.10)$$

In this equation, A is the gain used to convert the bit density μ into a reading in degrees Celsius. Note that μ lies between roughly $\frac{1}{3}$ and $\frac{2}{3}$ for the temperature range of interest. Using this equation, the output-referred noise of a given readout circuit can be estimated from the noise charge integrated during a single cycle in which ΔV_{BE} is integrated. This will be done later in this chapter for both CT and SC circuits. First, however, the theoretical noise limit imposed by the bipolar transistors will be determined.

5.1.4 Noise of the Bipolar Front-End

When evaluating the noise performance of a given readout circuit, it is essential to know how it compares to the limit imposed by the signal source, in this case the bipolar transistors that generate V_{BE} and ΔV_{BE} . This limit can be determined by calculating the noise charge $q_{n,\Delta V_{BE}}$ that would be accumulated by an ideal integrator during one clock period of the $\Sigma\Delta$ modulator as a result of the noise present on ΔV_{BE} . The output-referred noise σ_T can then be calculated using equation (5.10) derived in the previous section.

The noise produced by a diode-connected bipolar transistor is caused in part by the shot noise $i_{n,c}$ of its collector current, and in part by the thermal noise of its base resistance R_B [8]³. The resulting noise of the base-emitter voltage is

$$\begin{aligned} v_{n,V_{BE}}^2 &= \frac{i_{n,c}^2}{g_m^2} + 4kTR_B B = \left(\frac{kT}{qI}\right)^2 2qIB + 4kTR_B B \\ &= \frac{2kT}{g_m} B + 4kTR_B B, \end{aligned} \quad (5.11)$$

where g_m is the transistor's transconductance, and B is the noise bandwidth. For current levels in the μA range, $1/g_m$ is typically much larger than R_B , so that the base resistance term can be ignored.

³The contribution of $1/f$ noise is neglected in this analysis, not only because it is relatively small in bipolar transistors but also because it is eliminated when dynamic offset cancellation is employed to eliminate the mismatch between the transistors used for generating ΔV_{BE} .

The noise $i_{n,bias}$ of the bias current is translated to voltage noise via the transistor's impedance $1/g_m$. Assuming that $i_{n,bias}$ is dominated by the shot noise of the bias transistor⁴,

$$i_{n,bias}^2 = 2qIB, \quad (5.12)$$

the total noise of the base-emitter voltage is

$$v_{n,V_{BE}}^2 = \frac{2kT}{g_m}B + \frac{1}{g_m^2}i_{n,bias}^2 = \frac{4(kT)^2}{qI}B = \frac{4kT}{g_m}B, \quad (5.13)$$

that is, the base-emitter voltage is as noisy as a resistor of $1/g_m$.

The noise of the difference in base-emitter voltages ΔV_{BE} equals the squared sum of the noises of two base-emitter voltages. Since one of the transistors used in generating ΔV_{BE} carries a p times larger current than the other, it has a p times smaller noise power. The noise of the transistor with the smaller current therefore dominates:

$$v_{n,\Delta V_{BE}}^2 = 4\frac{(kT)^2}{qI} \left(1 + \frac{1}{p}\right)B \simeq \frac{4(kT)^2}{qI}B = \frac{4kT}{g_{m1}}B, \quad (5.14)$$

where I is the unit bias current, and g_{m1} the transconductance of the transistor carrying the smaller bias current⁵.

An ideal integrator will integrate $\alpha\Delta V_{BE}$ continuously during a clock period $t_{clk} = 1/f_{clk}$. This is equivalent to filtering it with a sinc filter, which has an effective noise bandwidth of $B = f_{clk}/2 = 1/(2 \cdot t_{clk})$ [9]. The rms noise charge accumulated in one clock period will therefore be

$$q_{n,\Delta V_{BE}} = \alpha \sqrt{\frac{2kT}{g_{m1}}t_{clk}}, \quad (5.15)$$

where the integrator gain is assumed to be one. The signal charge due to the integration of ΔV_{BE} and V_{BE} is $Q_{\Delta V_{BE}} = t_{clk} \cdot \alpha\Delta V_{BE}$ and $Q_{V_{BE}} = t_{clk} \cdot V_{BE}$, respectively. Substituting these values in (5.10) gives an output-referred noise of

$$\sigma_{T,bip} = \frac{A}{V_{REF}} \alpha \sqrt{\frac{2kT(1-\mu)}{g_{m1} \cdot N \cdot t_{clk}}}. \quad (5.16)$$

In this equation, A is the gain used to convert the bit density μ into a reading in degrees Celsius, N is the number of cycles of the $\Sigma\Delta$ modulator, and

⁴In practice, depending on the implementation of the bias circuit, the noise of the bias current may be larger than that given by (5.12).

⁵Note that any correlated noise in the two bias currents only results in a common-mode voltage noise, which does not contribute to $v_{n,\Delta V_{BE}}$. Such correlated noise comes, for instance, from noise sources in the circuit that generates the bias current.

$V_{REF} = V_{BE} + \alpha \Delta V_{BE}$. This shows that the noise limit imposed by the bipolar transistors can be reduced by increasing their g_m (which is achieved by increasing the bias current), by increasing the conversion time $t_{conv} = N \cdot t_{clk}$, or by reducing α (which implies using a larger current ratio p).

Using $A \simeq 600$ K, $V_{REF} \simeq 1.2$ V and $\mu = 0.5$, which is the bit density at $T = 300$ K, equation (5.16) can be written as

$$\sigma_{T,bip} \simeq (5.1 \cdot 10^{-9} \text{ K C}^{0.5}) \cdot \alpha \sqrt{\frac{1}{t_{conv} \cdot I}}. \quad (5.17)$$

If, for example, $t_{conv} = 100$ ms, $\alpha = 10$, and $I = 1 \mu\text{A}$, the output-referred noise will be $\sigma_T = 0.16$ mK. This shows that for most applications, the noise generated by the bipolar transistors is negligible. The output noise will usually be determined by noise introduced in the readout circuitry and/or by quantization noise.

5.2 Continuous-Time Circuitry

This section discusses the design of continuous-time (CT) readout circuitry for a smart temperature sensor. Several CT designs have been published [10–13]. In the designs of Bakker, which are described in detail in [14], a first-order $\Sigma\Delta$ modulator is used, based on the charge-balancing technique described in Section 4.1.3 (Figure 4.2b). This architecture will form the basis for the discussion in this section.

5.2.1 Implementation of Charge Balancing

The basic CT integrator shown in the introduction of this chapter (Figure 5.1b) cannot be directly connected to a bipolar transistor because it would draw bias current from the transistor, thus rendering its base-emitter voltage inaccurate. Instead, an integrator with a high input impedance is needed⁶. Two such integrators are shown in Figure 5.2.

Figure 5.2a shows a non-inverting integrator. A disadvantage of this integrator is that it introduces a zero at $\omega = 1/RC_{int}$. Moreover, it is difficult to implement charge balancing with it as its input would have to be switched back and forth between a positive and a negative voltage (e.g. $+\alpha \Delta V_{BE}$ and $-V_{BE}$). In a fully differential implementation, this could be realized, but then it would be complicated to implement a common-mode control that does not degrade the accuracy.

⁶Note that a switched-capacitor integrator also loads the bipolar transistors in its front-end. The resulting current drawn from the bipolar transistors is not a problem, however, as long as it is sufficiently small at the sampling instant at the end of a clock cycle. In contrast, in a CT integrator V_{BE} has to be accurate during the complete clock cycle.

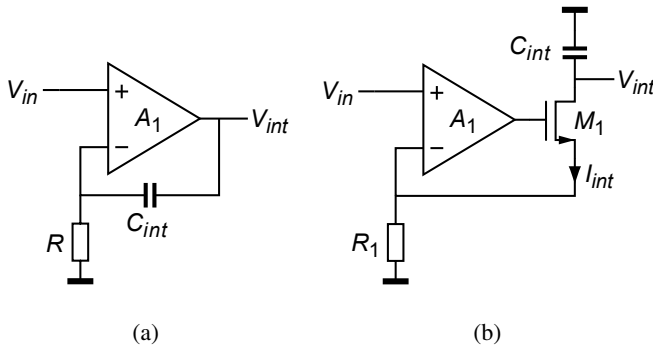


Figure 5.2. Continuous-time integrators with a high-impedance input: (a) non-inverting integrator, (b) integrator based on an active V-I converter.

A more attractive alternative is shown in Figure 5.2b: the input voltage is first converted into a current I_{int} using an active voltage-to-current (V-I) converter. This current is then integrated on the integration capacitor. Charge balancing can then be implemented using *two* such V-I converters that convert V_{BE} and ΔV_{BE} into currents with opposite polarity. This is the approach used by Bakker [14].

A CT $\Sigma\Delta$ modulator based on this idea is shown in Figure 5.3. Depending on the bitstream output bs , either a current proportional to ΔV_{BE} is applied to the integrator (when $bs = 0$), which is provided by a sinking V-I converter, or a current proportional to V_{BE} (when $bs = 1$), which is provided by a sourcing V-I converter. Since the V-I converters have a constant input voltage, their bandwidth (and hence power consumption) can be small. The use of an active integrator around opamp A_3 ensures that the outputs of the V-I converters see a virtual ground at V_B , rather than the voltage across the integration capacitor. Thus, their output-impedance requirements are relaxed. Moreover, their output currents can easily be switched to a low-impedance point at the same voltage V_B when they are not used, so that switching transients are minimized.

The circuit of Figure 5.3 offers two ways of implementing the gain α : a larger resistor can be used in the V-I converter for V_{BE} than in the V-I converter for ΔV_{BE} , and/or a shorter integration time can be used for V_{BE} than for ΔV_{BE} . If ΔV_{BE} is integrated during a full clock period $t_{clk} = 1/f_{clk}$, the integrated charge is

$$Q_{\Delta V_{BE}} = \frac{\Delta V_{BE}}{R_1} \cdot t_{clk}. \quad (5.18)$$

If V_{BE} is integrated during a fraction δ of a clock period, the associated charge is

$$Q_{V_{BE}} = \frac{V_{BE}}{R_2} \cdot \delta \cdot t_{clk}, \quad (5.19)$$

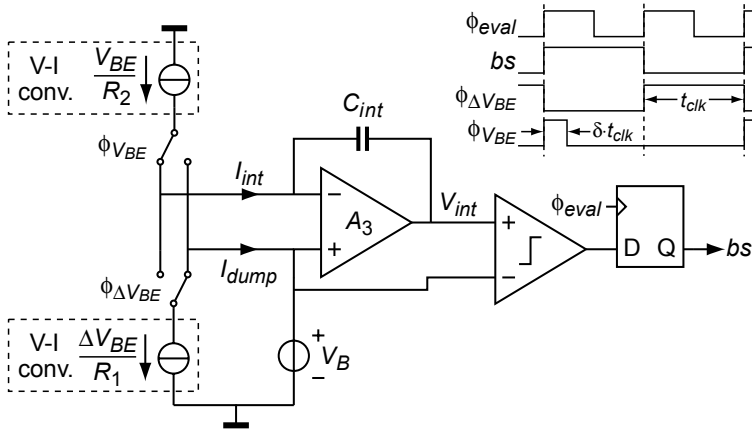


Figure 5.3. Continuous-time $\Sigma\Delta$ converter based on two V-I converters.

The gain α is therefore determined by the ratio of the resistors R_1 and R_2 , and by the fraction δ :

$$\alpha = \frac{R_2}{R_1 \cdot \delta}. \quad (5.20)$$

In the following, the accuracy and noise performance of the modulator of Figure 5.3 will be analyzed without going into too much implementation details. For now, the V-I converters are assumed to be implemented as shown in Figure 5.2b. Realizations of temperature sensors based on CT circuitry will be discussed in detail in Sections 7.1 and 7.2. There, specific examples will be given of how to interface the single-ended V-I converter of Figure 5.2b to the differential voltage ΔV_{BE} (Section 7.1.4), and how to produce a sourcing current from V_{BE} (Section 7.1.5). However, the general conclusions of the following sections are also applicable to other implementations.

5.2.2 Accuracy

Offset Errors

The input-referred offsets of the V-I converters add directly to V_{BE} and ΔV_{BE} . Since ΔV_{BE} is typically an order of magnitude smaller than V_{BE} , the offset requirement is most stringent for the V-I converter that has ΔV_{BE} as its input. Its maximum offset for a given temperature error ΔT can be found using (3.13):

$$|V_{os}| < \left(\frac{3}{\alpha} \text{ mV} / ^\circ\text{C} \right) \cdot \Delta T. \quad (5.21)$$

For $\Delta T = \pm 0.01^\circ\text{C}$ and $\alpha = 10$, for example, the maximum offset is $\pm 3 \mu\text{V}$. Given that CMOS opamps typically have offsets in the mV range, some form

of offset cancellation is required. An offset cancellation technique compatible with CT readout circuitry is chopping, which will be discussed in detail in Section 5.2.5.

The offset of opamp A_3 is much less critical. It affects the integrated currents via the finite output impedances $R_{out1,2}$ of the V-I converters, and is therefore attenuated by a factor R_{out1}/R_1 when referred back to the input of the sinking V-I converter, or by a factor R_{out2}/R_2 when referred back to the input of the sourcing V-I converter. Therefore, no special offset cancellation is needed for opamp A_3 .

Mismatch Errors

The accuracy of the gain α is determined by the matching of the resistors R_1 and R_2 , and by the accuracy with which the ratio of the integration times δ is defined. The latter can usually be done very accurately if the clock signals are derived from a synchronous counter, especially for the low clock frequencies typically used in temperature sensors (which are in the order of a few tens of kHz). The main concern in using a shorter integration time for V_{BE} is clock jitter rather than matching: for a given amount of clock jitter (i.e. a given uncertainty in the timing of the clock edges), the signal-to-noise ratio of $Q_{V_{BE}}$ decreases as δ is reduced.

Therefore, the accuracy of α is mainly determined by the matching of R_1 and R_2 . Matching in the order of 0.1% can be obtained if these resistors are constructed from identical unit elements arranged in a common-centroid pattern [15]. The minimum matching needed for a given temperature error ΔT can be found using (3.14):

$$\left| \frac{\alpha - \alpha_{ideal}}{\alpha_{ideal}} \right| < \left(\frac{2}{3} \% / ^\circ\text{C} \right) \cdot \Delta T, \quad (5.22)$$

which implies that for precision applications, the matching obtained from a precise layout is insufficient. Dynamic element matching can then be used to average out mismatch errors, as will be shown in Section 5.2.6.

Errors Due to Resistor Non-Linearity

The ratio of resistors R_1 and R_2 is not only determined by how well they are physically matched, but also by how well they are operated under identical bias conditions. This is the result of the voltage dependency of practical resistors. Since resistor R_1 is exposed to a much smaller voltage (ΔV_{BE}) than resistor R_2 (V_{BE}), this voltage dependency may result in a systematic error in the ratio R_2/R_1 , even if the resistors are constructed from perfectly matching unit elements. This error cannot be eliminated by means of dynamic element matching. Therefore, it is important to use resistors with a relatively small voltage dependency.

The voltage dependency is worst for diffused resistors, since their effective thickness depends on the width of the depletion-region of the reverse-biased junction that separates them electrically from the surrounding circuitry. The voltage dependency of p^+ resistors in an n-well can be reduced by placing each of the unit elements in its own n-well and tying this n-well to one side of the resistor. Thus, the resistor-well junction of all unit elements is biased in the same way. A better alternative is to use polysilicon resistors, which are generally much less voltage dependent [15].

Resistor non-linearity, as far as it is insensitive to processing spread, results in systematic non-linearity of the sensor. If this non-linearity is significant, it has to be taken into account in the design of the curvature-correction circuitry.

Errors due to Leakage Currents

Leakage currents are generated in reverse-biased junctions and can create an error in the integrated current. While such leakage currents are usually negligible at room temperature, they are highly temperature dependent and double approximately every eight degrees Celsius [15]. At 125 °C, they are typically in the order of 0.1 pA per μm^2 of junction area, and 0.1 pA per μm of junction periphery. The junctions of interest are the drain/source to bulk junctions of NMOS and PMOS transistors in the V-I converters and switches. If R_1 and R_2 are diffused resistors, their substrate or n-well junctions are also a major source of leakage, since the total area and periphery of these resistors may be significant. This is a second reason to use polysilicon resistors instead of diffused resistors (the first being their superior linearity).

Leakage acts as an offset on the integrated current. The maximum leakage current that can be tolerated can therefore be found by referring the leakage current back to the input of the V-I converters and comparing the resulting input-referred offset to equation (5.21). To obtain a maximum temperature error due to leakage in the order of 0.01 °C, for example, the maximum input-referred offset is in the order of $3 \mu\text{V}$ (see the paragraph above on offset errors). Suppose that $R_1 = 30 \text{ k}\Omega$. The maximum leakage current that corresponds to such an offset is then 0.1 nA, which translates into a maximum junction area in the order of $1000 \mu\text{m}^2$.

Errors due to Finite Gain

Finite gain introduces errors both in the V-I converters and in the integrator. In the V-I converters, finite loop gain results in errors in the closed-loop transconductances. Ideally, these transconductances are equal to $1/R_1$ and $1/R_2$, but finite gain introduces an error inversely proportional to the loop gain $A_{OL,VI}$. Since the loop gains in the two V-I converters cannot be expected to match, the resulting errors in both converters should be reduced to negligible levels. As these errors are equivalent to an error in α , the required loop gain

can be derived from (5.22):

$$A_{OL,VI} > \frac{1}{\left(\frac{2}{3}\%/\text{°C}\right) \cdot \Delta T} = \frac{150\text{ °C}}{\Delta T}. \quad (5.23)$$

For an error contribution of 0.01 °C, a loop gain of 15000, or 84 dB, is required.

Finite gain of opamp A_3 results in a non-zero overdrive voltage at its input, which modulates the current I_{int} due to the finite output impedances of the V-I converters. Assuming the opamp is implemented as a transconductance amplifier, there are two main causes of this non-zero overdrive voltage [2]. The first cause is the finite transconductance g_{m3} of the opamp, which implies that an overdrive voltage is required to provide the feedback current. This overdrive voltage becomes very large if the opamp is slewing. This condition should be avoided by ensuring that the maximum current that the output of the opamp can provide is well above the maximum current I_{int} supplied by the V-I converters. If the opamp is not slewing, the overdrive is equal to I_{int}/g_{m3} . The change in the integrated current is then

$$\frac{\Delta I_{int}}{I_{int}} \simeq \frac{1}{g_{m3}R_{out}}, \quad (5.24)$$

where R_{out} is the output impedance of either the sinking or the sourcing V-I converter. Since these output impedances cannot be expected to match, the change in the integrated current has to be reduced to negligible levels. Since a relative error in the integrated current is equivalent to an error in α , the product $g_{m3}R_{out}$ has to meet the same requirement as the loop gain of the V-I converters:

$$g_{m3}R_{out} > \frac{150\text{ °C}}{\Delta T}. \quad (5.25)$$

The second cause of a non-zero overdrive voltage at the input of A_3 is its finite DC gain $A_{0,3}$, which implies that an overdrive voltage is required to produce the output voltage V_{int} . As a result, a fraction of V_{int} appears at the input of the opamp, which in turn modulates the integrated current due the finite output impedances of the V-I converters:

$$\frac{\Delta I_{int}}{I_{int}} = \frac{V_{int}}{A_{0,3}R_{out}}, \quad (5.26)$$

where R_{out} is again the output impedance of either V-I converter. As discussed in Sections 4.3.4 and 4.4.6, such leakage of the integrator's output voltage to its input limits the ENOB of the $\Sigma\Delta$ modulator. If the sinking current source is connected to the integrator, the effective DC gain from the input of the V-I converter to the output of the integrator is

$$A_0 = \frac{A_{0,3}R_{out1}}{R_1}, \quad (5.27)$$

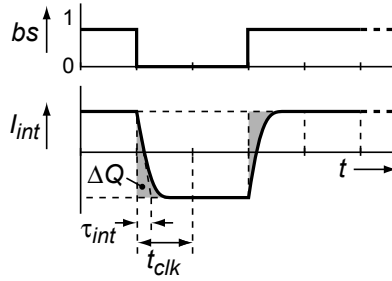


Figure 5.4. Switching transients in the integrated current I_{int} lead to inter-symbol interference.

where R_{out1} is the output impedance of the sinking V-I converter. The minimum required DC gain can then be derived from equations (4.24) and (4.30) for a first- and a second-order modulator, respectively. The DC gain A_0 should also be large enough to ensure that errors and noise introduced after the first integrator are negligible when referred back to the input.

Errors due to Switching Transients

When the outputs of the V-I converters are switched, small transients will occur in the integrated current, resulting in an error in the integrated charge. These transients are mainly due to the finite bandwidth of the integrator around opamp A_3 , which is caused by its load capacitance C_L and any parasitic capacitance C_i at its input. Assuming that these capacitances are much smaller than the integration capacitor C_{int} and that the opamp can be modelled as a transconductance g_{m3} , the integrated current will change exponentially in response to a step in the input current, with a time constant of

$$\tau_{int} = \frac{C_i + C_L}{g_{m3}}. \quad (5.28)$$

This is illustrated in Figure 5.4. When the bitstream changes from one to zero, I_{int} changes from V_{BE}/R_2 to $-\Delta V_{BE}/R_1$. The charge integrated during the following clock period then deviates from the ideal value given by equation (5.18) by an amount of

$$\begin{aligned} \Delta Q &\simeq \left(\frac{\Delta V_{BE}}{R_1} + \frac{V_{BE}}{R_2} \right) \frac{\tau_{int}}{2} \\ &= (Q_{\Delta V_{BE}} + Q_{V_{BE}}) \frac{\tau_{int}}{2t_{clk}} \quad (\tau_{int} \ll t_{clk}). \end{aligned} \quad (5.29)$$

If the time constants of rising and falling transients are the same, the same charge error occurs when the bitstream changes from zero to one.

As shown in Figure 5.4, these errors only occur on transitions in the bitstream, and thus make the integrated charge in a given clock cycle dependent on the bitstream value in the previous cycle. This so-called inter-symbol interference introduces extra non-linearity and increases the quantization noise of the $\Sigma\Delta$ modulator [1, 2]. It can be prevented by using return-to-zero (RTZ) switching, which means that I_{int} is switched to zero at the end of every clock cycle by switching off the outputs of both V-I converters. Thus, every clock cycle contains a rising and a falling transient, and dependency on the previous cycle is eliminated.

Even if RTZ switching is used, a residual error is caused by mismatch $\Delta\tau_{int}$ between the time constants associated with the integration of V_{BE} and that of ΔV_{BE} . The resulting charge errors are equivalent to an error in α of $\Delta\tau_{int}/t_{clk}$. The maximum $\Delta\tau_{int}$ can therefore be found using (5.22).

Switching transients may also result from the use of dynamic error correction techniques in the V-I converters or in the front-end circuitry. For instance, if the dynamic element matching (DEM) scheme of Figure 3.8 is applied to generated ΔV_{BE} , the output of the sinking V-I converter will contain DEM residuals, which have to be averaged out. Similarly, if chopping is applied to reduce the offset of the V-I converters (see Section 5.2.5), the output of the V-I converters will contain a modulated offset, which also has to be averaged out. Care has to be taken to arrange the timing of the integrator in such a way that the integrated charge indeed corresponds to the desired average. It may be helpful to briefly interrupt the integration during the switching in the V-I converters or in the front-end circuitry, so as to prevent switching transients from being integrated [16, 17].

5.2.3 Noise

Using the per-cycle analysis introduced in Section 5.1.3, the noise performance of the modulator of Figure 5.3 can be derived from the noise charge integrated during a single $\Sigma\Delta$ cycle in which ΔV_{BE} is integrated. This noise charge is determined by the noise introduced in the V-I converter (Figure 5.2b). The dominant noise sources in this V-I converter are the noise voltage of the opamp and that of resistor R_1 . These noise voltages add to the noise present in the input voltage ΔV_{BE} , which is given by equation (5.14). This leads to a total input-referred noise of

$$v_{n,in}^2 = v_{n,\Delta V_{BE}}^2 + v_{n,R_1}^2 + v_{n,opamp}^2 = 4kT \cdot R_{n,CT} \cdot B, \quad (5.30)$$

where B is the noise bandwidth, and $R_{n,CT}$ is the equivalent input noise resistance of the circuit:

$$R_{n,CT} = \frac{1}{g_{m1}} + R_1 + \frac{\gamma}{g_{m,opamp}}. \quad (5.31)$$

In this equation, g_{m1} is the smallest transconductance of the pair of bipolar transistors used to generate ΔV_{BE} , and $g_{m,opamp}$ is the transconductance of the input pair of the opamp. The factor γ accounts for the noise contributions of the various signal transistors in the opamp. If the noise is dominated by the input pair, γ is approximately $4/3$ [18]. The opamp's $1/f$ noise is neglected in this analysis, because it can be eliminated by means of dynamic offset cancellation (see Section 5.2.5).

The input-referred noise voltage leads to a noise current, which is integrated for the duration of a clock period t_{clk} of the $\Sigma\Delta$ modulator. This is equivalent to filtering the noise voltage with a sinc filter, which has an effective noise bandwidth of $B = f_{clk}/2 = 1/(2 \cdot t_{clk})$ [9]⁷. The integrated noise charge is therefore

$$q_{n,\Delta V_{BE}} = \frac{v_{n,in}}{R_1} \cdot t_{clk} = \frac{1}{R_1} \sqrt{2kT \cdot R_{n,CT} \cdot t_{clk}}. \quad (5.32)$$

Substituting this value along with (5.18) and (5.19) in equation (5.10) yields the output-referred temperature noise:

$$\begin{aligned} \sigma_{T,CT} &= A \cdot \frac{\alpha \sqrt{2kT \cdot R_{n,CT} \cdot t_{clk}}}{\alpha \Delta V_{BE} \cdot t_{clk} + V_{BE} \cdot t_{clk}} \sqrt{\frac{(1-\mu)}{N}} \\ &= \frac{A}{V_{REF}} \alpha \sqrt{\frac{2kT \cdot R_{n,CT} \cdot (1-\mu)}{N \cdot t_{clk}}}. \end{aligned} \quad (5.33)$$

This shows that the output noise can be reduced by increasing the conversion time $t_{conv} = N \cdot t_{clk}$, or by reducing the equivalent input noise resistance $R_{n,CT}$. Depending on which noise source is dominant, the latter can be done by increasing the bias currents of the bipolar transistors, by increasing the bias current of the opamp, or by reducing R_1 .

Using $A \simeq 600$ K, $V_{REF} \simeq 1.2$ V and $\mu = 0.5$, which is the bit density at $T = 300$ K, equation (5.33) can be written as

$$\sigma_{T,CT} \simeq (32 \cdot 10^{-9} \text{ K } \Omega^{-0.5} \text{ s}^{0.5}) \cdot \alpha \sqrt{\frac{R_{n,CT}}{t_{conv}}}. \quad (5.34)$$

This equation can be used to calculate the noise resistance required for a given output-referred noise. If, for example, $\alpha = 10$, $t_{conv} = 100$ ms, and an output-referred noise of $\sigma_{T,CT} = 1$ mK is desired, $R_{n,CT}$ should be less than 970 k Ω . If $R_1 = 60$ k Ω and $1/g_{m1} = 25$ k Ω (which corresponds to a bias current of 1 μ A), a noise budget of 890 k Ω remains for the opamp.

⁷Although the V-I converter limits the bandwidth of the noise current, its corner frequency is assumed to be significantly larger than B , so that the noise current can be assumed to be white over the bandwidth of interest.

By dividing equation (5.33) by equation (5.16), the noise performance of the CT circuit can be compared to that of just the bipolar front-end:

$$\frac{\sigma_{T,CT}}{\sigma_{T,bip}} = \sqrt{R_{n,CT} \cdot g_{m1}}. \quad (5.35)$$

For the above example above, the noise of the CT circuit is about 6 times larger than that of the front-end.

Noise due to clock jitter has been ignored in the above analysis. Clock jitter introduces an uncertainty in the clock period t_{clk} , which results in additional noise charge. With the relatively low clock frequencies used in temperature sensors, it is usually not a problem to reduce this additional noise to negligible levels.

5.2.4 Power Consumption

While it is not the main goal of this work to minimize power consumption, it is important to verify whether a given readout circuit can provide the required performance at a power consumption that does not cause significant self-heating. From Section 5.1.2, it can be concluded that self-heating is negligible if the average supply current is in the order of at most a few tens of μA .

The supply current of a CT implementation is usually not determined by noise requirements [14]. From the noise analysis in the previous section it is clear that even for very low noise requirements (e.g. 1 mK), a fairly large noise resistance can be tolerated ($\sim 1\text{ M}\Omega$). This implies that very low bias currents can be used (well below $1\ \mu\text{A}$). This is the result of the large conversion time, which leads to a small effective noise bandwidth.

Rather than by noise requirements, the supply current will be dictated by the accuracy requirements discussed in Section 5.2.2. For instance, the bias current of the integrator has to be large enough to prevent slewing, to ensure negligible overdrive at its input, and to guarantee short enough settling transients. The overall power consumption will strongly depend on the values of the resistors R_1 and R_2 , which should be chosen such that the integrated current is large enough compared to leakage currents. As a result, the supply current will be highly implementation dependent.

If the supply current needed to meet the accuracy requirements results in a noise much lower than required, it can be worthwhile to reduce the conversion time, so that the sensor can be powered down part of the time while maintaining the same output data rate f_{out} . So far, it has been assumed that the sensor is continuously powered, i.e. that the conversion time t_{conv} is equal $1/f_{out}$. If the sensor is powered only a fraction x of the time, the conversion time is reduced to

$$t_{conv} = \frac{x}{f_{out}}, \quad (0 < x < 1). \quad (5.36)$$

To maintain the same number of $\Sigma\Delta$ cycles, the clock frequency has to be increased by the same factor:

$$f_{clk} = \frac{N}{t_{conv}} = \frac{N}{x} f_{out}. \quad (5.37)$$

To maintain the same accuracy at this higher clock frequency, the supply current will have to be increased to ensure that errors related to switching transients remain the same. The supply current will typically be composed of a constant part and a part roughly proportional to the clock frequency:

$$I_{supply} = I_{const} + K \cdot f_{clk}. \quad (5.38)$$

The constant part includes the bias currents of the V-I converters, the bipolar transistors, and support circuitry, such as the bias circuit and the oscillator. The frequency-dependent part includes the supply current of the integrator. Combining the above equations, the average power consumption can be written as

$$P_{diss} = x \cdot V_{DD} \cdot \left(I_{const} + K \frac{N}{x} f_{out} \right) \quad (5.39)$$

$$= V_{DD} \cdot (x \cdot I_{const} + K \cdot N \cdot f_{out}). \quad (5.40)$$

This shows that the constant part can be reduced by reducing the duty cycle x down to the point where the noise is no longer negligible.

5.2.5 Chopping

As mentioned in Section 5.2.2, offset is a major source of inaccuracy in both CT and SC readout circuits. Offsets in the μV range are required, while CMOS amplifiers typically have offsets in the mV range. Offset is caused by transistor mismatch (for instance mismatch in the input pair of an amplifier). While this mismatch can be reduced by using transistors with a larger area [19], the offset levels required in a temperature sensor cannot be obtained with practical transistor sizes. Trimming could be used to reduce the offset, but it cannot compensate for offset drift and temperature dependency⁸ [14].

The solution to this problem lies in dynamic offset cancellation techniques [20]. These techniques can be categorized in two groups:

- *Autozeroing techniques*: offset is sampled and then subtracted from the input.

⁸This is because the offset of a CMOS differential pair is caused by mismatch of both the threshold voltage V_T and the current factor β , each of which are caused by various independent physical phenomena [19]. As a result, the temperature dependency of the offset spreads. Trimming therefore cannot be used if a low offset is required over the full operating temperature range. This is an important difference with bipolar amplifiers, whose offset has a well-defined PTAT characteristic and can therefore be trimmed successfully [8].

- *Chopping techniques:* offset is modulated away from the signal band and then filtered out.

Both autozeroing and chopping are dynamic processes that remove the offset during operation of the amplifier. As a result, in contrast to trimming, they also eliminate offset drift and $1/f$ noise, both of which can be seen as slowly varying offsets.

Autozeroing, being a sampled-data technique, naturally complements SC integrators and will be discussed in more detail in Section 5.3.5. In principle, autozeroing could also be used to reduce the offset in a CT circuit. In an autozeroed amplifier, however, noise is sampled along with the offset. As a result, wide-band noise is undersampled and aliases to the baseband, resulting in an increased noise floor. In a SC integrator, the input signal is sampled anyway, which makes the aliasing of wide-band noise inevitable. One of the more attractive features of a CT integrator, however, is that this aliasing of noise does not occur. Autozeroing is therefore not an attractive technique to reduce the offset in a CT implementation.

Chopper Amplifier

A dynamic offset cancellation technique more suitable for CT circuits is chopping. The principle of a chopper amplifier is shown in Figure 5.5 [14]. The amplifier's input voltage V_{in} is first passed through a polarity-reversal switch driven by a clock signal ϕ_{ch} . This 'chopper switch' (see Figure 5.5b) periodically reverses the polarity of the input signal and thus modulates it by a square wave (Figure 5.5c). The modulated input signal is passed through a differential amplifier A with an offset V_{os} . At the output of this amplifier, the amplified input signal is therefore found at the harmonics of ϕ_{ch} , while the amplified offset is found at DC (see Figure 5.5d). A second chopper switch demodulates the amplified input signal back to DC, and at the same time modulates the offset to the harmonics of ϕ_{ch} , where they are filtered out by a low-pass filter (LPF). What remains is the amplified input signal without offset.

Low-frequency signals such as drift and $1/f$ noise will be modulated and filtered out along with the offset. Like autozeroing, chopping therefore also solves the problem of the relatively large $1/f$ noise of CMOS amplifiers. In contrast with autozeroing, chopping is a modulation technique rather than a sampling technique. As a result, no undersampling of wide-band noise occurs, and the input-referred noise of the amplifier is the same as its thermal noise.

Chopped Voltage-to-Current Converter

Figure 5.6 shows how a chopper amplifier can be incorporated in the V-I converter of Figure 5.2b [14]. The opamp is split up into two stages: a chopped fully differential first stage with gain A_{11} and a second stage with gain

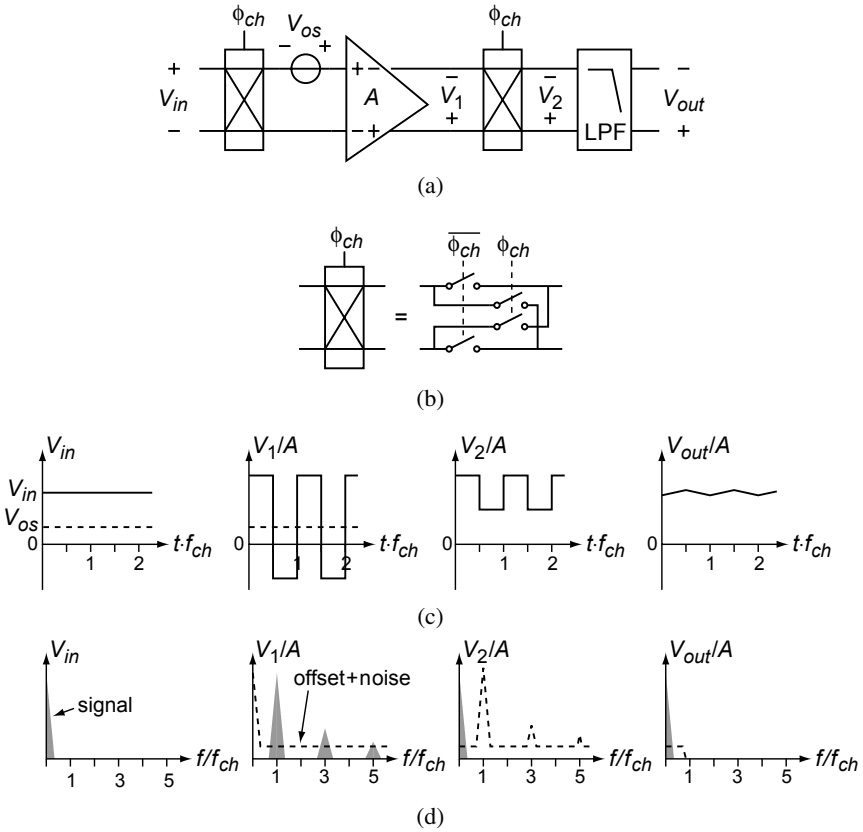


Figure 5.5. Offset cancellation using the chopping technique: (a) a chopper amplifier; (b) implementation of a chopper switch; (c) voltages in the amplifier as a function of time; (d) frequency spectrum of the same voltages.

A_{12} , which takes care of the differential-to-single-ended conversion needed to drive the gate of M_1 . The chopping effectively inverts the offset V_{os1} of the first stage. As a result, the output current switches back and forth between $(V_{in} \pm V_{os1}) / R_1$ (ignoring, for now, the offset of the second stage). Provided that the duty cycle of the control signal ϕ_{ch} is exactly 50%, the larger current is integrated during the same amount of time as the smaller current, and the offset component is averaged out. At the end of every clock cycle, the integrator's output is offset-free. Thus, no separate low-pass filter is needed.

The offset V_{os2} of the second stage is not chopped. To refer this offset back to the input, it needs to be divided by the gain of the first stage at the chopping frequency f_{ch} . Note that this gain is less than the DC gain of the first stage if f_{ch} is larger than its corner frequency. The first stage therefore needs to have a high enough DC gain *and* a high enough bandwidth to reduce the input-referred

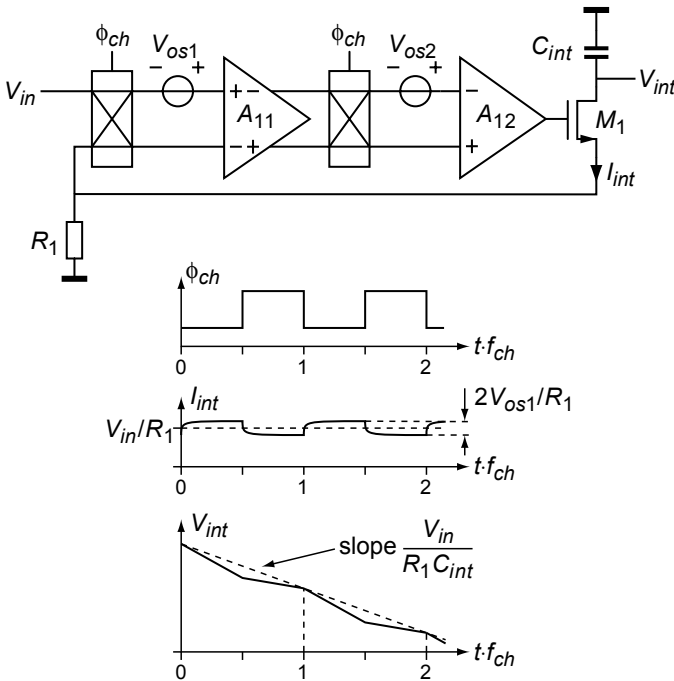


Figure 5.6. Chopped continuous-time integrator, with chopper control signal ϕ_{ch} , integrated current I_{int} , and output voltage V_{int} as a function of time.

offset due to the second stage to negligible levels. An efficient way of ensuring this is to use an OTA as first stage (e.g. a folded-cascode stage) and a Miller-compensated second stage [14]. The unity-gain bandwidth of the complete amplifier is then determined by the g_m of the first stage and the Miller capacitor. The unity-gain bandwidth of the chopped first stage, however, is much larger, as it is determined by the same g_m and the (parasitic) load capacitance at the output of the first stage.

Residual Offset of Chopper Amplifiers

The residual offset of chopper amplifiers is usually determined by charge injection and clock feed-through in the chopper switches [20]. These phenomena give rise to voltage spikes. As far as these spikes occur at the input or output of the amplifier, as shown in Figure 5.7, they will be demodulated by the second chopper and result in an average DC offset. The charge Q_{inj1} injected by the input chopper switch is usually dominant, although the offset due to the charge Q_{inj2} injected by the output switch may not be negligible if the amplifier has a high-impedance output.

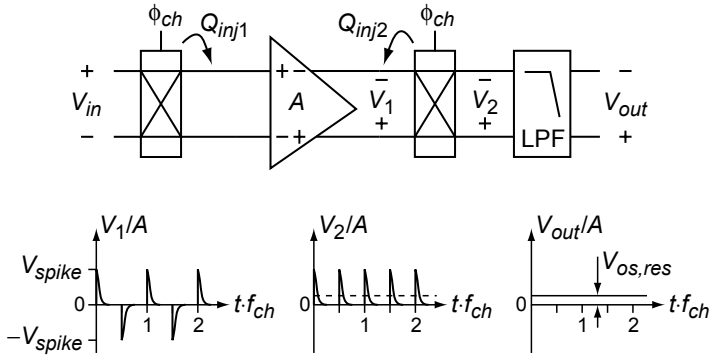


Figure 5.7. Residual offset of a chopper amplifier as a result of demodulation of charge-injection spikes.

If the time constant of the spikes τ_{spike} is much smaller than the period $1/f_{ch}$ of the control signal ϕ_{ch} , the input-referred residual offset is approximately equal to

$$V_{os,res} = 2\tau_{spike} f_{ch} V_{spike}, \quad (5.41)$$

where V_{spike} is the peak amplitude of the spikes [20]. This amplitude depends on the amount of injected charge and on the impedance at the node where the charge is injected. The amount of charge depends, in turn, on the size of the switches, and the amplitude and slew rate of the clock signal (see also Section 5.3.5). Since chopper switches are *differential* circuits, the spikes are ideally common-mode signals that will not result in a differential-mode offset. In practice, however, differential spikes result from unbalance in the impedances and mismatch between the switches. In the circuit of Figure 5.6, for example, the unmatched impedances at the input of the amplifier will lead to differential spikes.

As shown by (5.41), the residual offset is proportional to f_{ch} . An obvious way to reduce the offset is therefore to reduce f_{ch} . This is limited by the corner frequency of the $1/f$ noise. If f_{ch} is chosen lower than this corner frequency, the $1/f$ noise will not be completely modulated away from DC, so that the input-referred noise at DC will be higher than the thermal noise.

The residual offset of chopper amplifiers is typically in the order of a few tens of μV . Since this is an order of magnitude larger than needed for processing of ΔV_{BE} with negligible errors, additional techniques are required. Such advanced offset cancellation techniques will be discussed in Section 5.4.

5.2.6 Dynamic Element Matching

As was mentioned in Section 5.2.2, the accuracy of the gain α is mainly determined by the accuracy of the ratio of the resistors R_1 and R_2 in the V-I converters (see Figure 5.3). As a result of mismatch, this accuracy is limited to typically 0.1% [15], while an accuracy in the order of 0.01% is desired.

If the resistors R_1 and R_2 are constructed from nominally equal unit elements, dynamic element matching (DEM) can be used to reduce mismatch errors [21]. As discussed in Section 3.2.2, this means that the unit elements are interchanged in a number of steps. In each of the individual steps, there will be a mismatch error, but these errors average out over all steps.

When applying DEM in a CT circuit, two potential problems have to be taken into account. First of all, DEM introduces extra switching transients. If these transients are integrated, they are likely to introduce errors. Such errors can be avoided if the output of the V-I converter in question is disconnected from the integrator during the switching [17].

The second problem is related to the non-zero on-resistances of the switches used to interchange the unit elements. The switches have to be designed in such a way that voltage drop across their on-resistance does not introduce significant errors. This can be realized by using wide switches, which have a low on-resistance, but such switches require a large chip area and inject a lot of charge when they switch.

A more attractive alternative is to use so-called Kelvin connections, i.e. connections consisting of a force and a sense line [22]. The force line is driven by a (high-ohmic) current source and can therefore contain an arbitrary switch resistance. The sense line does not carry any current and therefore this line too can contain an arbitrary switch resistance. Ideally, all resistors that have to be interchangeable in a DEM scheme should have terminals that are either shared by all resistors (so that no switches are needed) or connected via a Kelvin connection (so that switches can be used).

Figure 5.8 shows how the resistors of a sinking and a sourcing V-I converter can be dynamically matched using Kelvin connections. Two nominally equal resistors $R_1 = R(1 + \delta)$ and $R_2 = R(1 - \delta)$ are used, which are both connected to ground (a shared terminal). The switches used are either connected to an opamp input (sense lines), or to the MOS transistors (force lines). In the sourcing V-I converter (around opamp A_2), an extra degree of freedom in the scaling of the output current is obtained by using a $k : 1$ ratio in the PMOS transistors. The gain α thus becomes

$$\alpha = \frac{R_2}{R_1} \frac{k}{\delta}, \quad (5.42)$$

where, as before, δ is the ratio between the integration times of the sinking and the sourcing current.

An important problem of the circuit in Figure 5.8 is that ΔV_{BE} is not available as a single-ended voltage (that is, as a voltage referenced to ground). Because it is generated as the difference in base-emitter voltages of two diode-connected pnp transistors, it is offset from ground by a common-mode voltage equal to a base-emitter voltage. A sinking V-I converter that can handle this common-mode voltage will be discussed in Section 7.1.4. However, that implementation is not compatible with the DEM scheme of Figure 5.8, as it does not have a pure Kelvin connection to resistor R_1 . A fully differential version of the circuit in Figure 5.8 could be used to solve this problem. But because the above-mentioned problems can be much more conveniently avoided by using switched-capacitor readout circuitry, as will be shown in the next section, this alternative will not be investigated further.

5.3 Switched-Capacitor Circuitry

This section focuses on the implementation of smart temperature sensors using switched-capacitor (SC) techniques. While the majority of the designs found in literature are continuous-time designs, some SC designs have been reported [23, 24]. In this section, the performance of a $\Sigma\Delta$ modulator based on a basic SC integrator is analyzed, again using the charge balancing scheme of Figure 4.2b.

5.3.1 Implementation of Charge Balancing

In a SC integrator (see Figure 5.1a), the input voltage is sampled on a capacitor C_S and the resulting charge is transferred to an integration capacitor. The charge balancing between V_{BE} and ΔV_{BE} can therefore be realized by sampling either $-V_{BE}$ or ΔV_{BE} in a given clock cycle of the $\Sigma\Delta$ modulator (depending on the modulator's bitstream).

Figure 5.9 shows how the circuit of Figure 5.1a can be modified to integrate V_{BE} . In phase ϕ_1 , the base-emitter voltage V_{BE} of a diode-connected substrate pnp transistor, biased at a current I , is sampled on capacitor C_S . The resulting charge on this capacitor is

$$Q_{V_{BE}} = C_S \cdot V_{BE}. \quad (5.44)$$

In phase ϕ_2 , this charge is transferred to C_{int} .

Figure 5.10 shows how the circuit can be reconfigured to integrate ΔV_{BE} . During phase ϕ_1 , a current I is passed through the bipolar transistor, so that $V_{BE}(I)$ is sampled. In phase ϕ_2 , the bias current is increased to a total of pI . As a result, the voltage across the sampling capacitor increases by ΔV_{BE} . The charge required for this increase is accumulated on C_{int} . This is repeated N_α times, i.e. one $\Sigma\Delta$ cycle consists of N_α charge transfers. The total charge

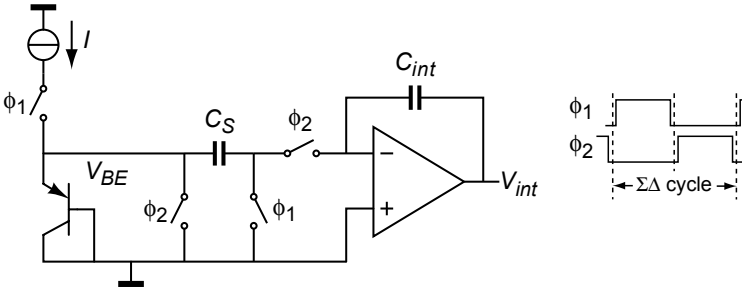


Figure 5.9. Integration of V_{BE} using a switched-capacitor integrator.

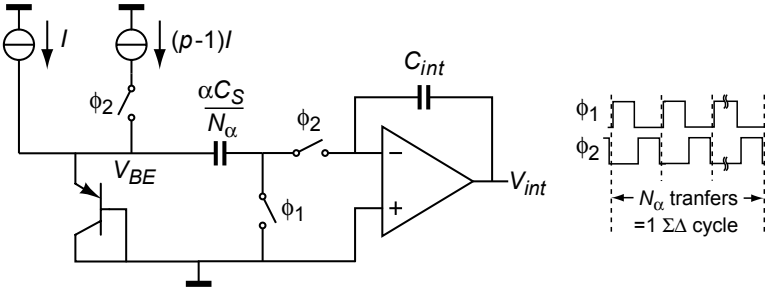


Figure 5.10. Integration of ΔV_{BE} using a switched-capacitor integrator.

transferred in one $\Sigma\Delta$ cycle is therefore

$$Q_{\Delta V_{BE}} = N_{\alpha} \cdot \frac{\alpha C_S}{N_{\alpha}} \cdot \Delta V_{BE} = \alpha C_S \cdot \Delta V_{BE}. \quad (5.45)$$

The gain α is can thus be realized by using a larger sampling capacitor for ΔV_{BE} than for V_{BE} , and/or by performing multiple charge transfers in one cycle of the $\Sigma\Delta$ modulator. If a single charge transfer is used for ΔV_{BE} , the sampling capacitor size for ΔV_{BE} has to be αC_S . If N_{α} charge transfers are used, the sampling capacitor can be a factor N_{α} smaller. The pros and cons of these alternatives will be investigated in the following sections.

In practice, the circuits of Figure 5.9 and Figure 5.10 would be merged to form a $\Sigma\Delta$ modulator that can either integrate V_{BE} or ΔV_{BE} based on its bitstream. A detailed example of such a modulator will be presented in Section 7.3. In the rest of this section, the sub-circuit of Figure 5.10, which integrates ΔV_{BE} , will be analyzed further. As shown in Section 5.1.3, the accuracy and noise of this circuit can be used to estimate the overall performance of a SC $\Sigma\Delta$ modulator.

5.3.2 Accuracy

Offset Errors

The accuracy of the circuit of Figure 5.10 is determined by various non-idealities. An important non-ideality is the offset V_{os} of the opamp. This adds directly to ΔV_{BE} . As in a CT implementation, the maximum offset for a given temperature error ΔT can therefore be found using (3.13):

$$|V_{os}| < \left(\frac{3}{\alpha} \text{ mV} / ^\circ\text{C} \right) \cdot \Delta T. \quad (5.46)$$

For $\Delta T = \pm 0.01^\circ\text{C}$ and $\alpha = 10$, for example, the maximum offset is $\pm 3 \mu\text{V}$. Given that CMOS opamps typically have offsets in the mV range, some form of offset cancellation is clearly required.

Autozeroing is a suitable offset-reduction technique for use in SC circuits and will be discussed in Section 5.3.5. As will be shown there, the residual offset in an autozeroed integrator is usually determined by the charge injected by the switches onto the sampling capacitor and the integration capacitor. It is therefore inversely proportional to the capacitor size, and proportional to the number of charge transfers N_α (since charge is injected at the end of every clock phase). In terms of offset, it is therefore desirable to minimize N_α .

Mismatch Errors

The accuracy of the gain α is determined by the matching between the sampling capacitor $\alpha C_S / N_\alpha$ used for ΔV_{BE} and the capacitor C_S used for V_{BE} . If $N_\alpha = \alpha$, the same capacitor can be used, so that mismatch errors are eliminated. The gain is then completely realized by means of multiple charge transfers. However, as mentioned above, the choice $N_\alpha = \alpha$ leads to larger charge-injection-related errors.

If $N_\alpha = 1$, the gain is realized using a single charge transfer with an α times larger sampling capacitor. Matching in the order of 0.1% can then be obtained if this capacitor is constructed from a parallel combination of α capacitors that have the same layout as capacitor C_S , and are organized in a common-centroid pattern [15]. The minimum matching needed for a given temperature error ΔT can be found using (3.14):

$$\left| \frac{\alpha - \alpha_{ideal}}{\alpha_{ideal}} \right| < \left(\frac{2}{3} \% / ^\circ\text{C} \right) \cdot \Delta T, \quad (5.47)$$

which implies that, for precision applications, the matching obtained from a precise layout is insufficient. Dynamic element matching can then be used to average out mismatch errors, as will be shown in Section 5.3.6.

Errors due to Capacitor Non-Linearity

Just as a CT implementation is affected by voltage dependency of resistors, a SC implementation suffers from voltage dependency of the sampling capacitors. This disturbs the sampling-capacitor ratio, an effect that cannot be eliminated by means of dynamic element matching. The resulting non-linearity, as far as it is insensitive to processing spread, can be treated as a systematic error and corrected for by taking it into account in the design the curvature correction circuitry.

In standard (digital) CMOS, two flavours of capacitors are available: MOS capacitors and metal sandwich capacitors. Metal sandwich capacitors are linear but they have a very small capacitance per area (e.g. $0.1 \text{ fF}/\mu\text{m}^2$) and are therefore usually too big for use in temperature sensors. MOS capacitors have a much larger capacitance per area (in the order of $2 \text{ fF}/\mu\text{m}^2$). Their voltage dependency, however, is much worse: their capacitance can change up to a factor of five as a function of the applied voltage [25]. The resulting non-linearity cannot be expected to be purely systematic and is beyond what can be corrected for using curvature-correction techniques.

In general, even-order non-linearity can be reduced by using fully differential circuitry. Moreover, several techniques have been published for constructing more linear capacitors from a combination of MOS capacitors [25, 26]. Linearities better than 10 bits have been reported, which reduces the resulting non-linearity to the same order of magnitude as that due to the curvature of V_{BE} . It is unclear whether this non-linearity is reproducible enough to be considered a systematic error. Some processes offer linearized MOS capacitors or poly-poly capacitors as an option. These devices offer a capacitance per area that is comparable to that of regular MOS capacitors, and they are usually so linear that they do not introduce significant errors. However, the associated extra processing steps imply an increase in production costs.

Errors due to Leakage Currents

Leakage currents (see also Section 5.2.2) result in droop of the voltage stored on the integration capacitor. If the total leakage current is I_{leak} , the charge lost in one cycle equals

$$\Delta Q = I_{leak} \cdot t_{clk} = \frac{I_{leak}}{f_{clk}}. \quad (5.48)$$

This shows that errors due to leakage currents are worse for lower clock frequencies. Leakage results mainly from reverse-biased pn-junctions, which generate a leakage current that increases rapidly with temperature and is in the order of 0.1 pA per μm^2 of junction area at 125°C . Examples of such junctions are the source-bulk and drain-bulk junctions of the switches, as well as the well-substrate junction of MOS capacitors. For this reason, double poly capacitors, if available, are preferred over (linearized) MOS capacitors.

The charge ΔQ can be translated into an input-referred offset voltage by dividing it by the sampling capacitance. The maximum leakage can then be found by comparing this input-referred offset voltage to the maximum offset given by (5.46). To obtain a maximum temperature error due to leakage in the order of 0.01°C , for example, the maximum input-referred offset is in the order of $3\ \mu\text{V}$. Suppose that $f_{clk} = 10\ \text{kHz}$ and the sampling capacitance $\alpha C_S = 100\ \text{pF}$. The maximum leakage current is then $3\ \text{pA}$, which corresponds to a maximum junction area in the order of $30\ \mu\text{m}^2$.

In a fully differential implementation, leakage will be mainly a common-mode effect. In that case, the *mismatch* between the leakage currents of the differential half-circuits is of interest, which can be assumed to be an order of magnitude smaller than the absolute leakage current. As a result, the junction area can be an order of magnitude larger.

Settling Errors

Another non-ideality is incomplete settling of the voltage across the sampling capacitor. If this settling were linear, the relative settling error in $Q_{\Delta V_{BE}}$ would be the same as that in $Q_{V_{BE}}$ and would thus cancel. The settling, however, is highly non-linear, due to slewing and the non-linearity of the bipolar transistor. The circuit should therefore be designed such that settling errors are negligible. Since a relative settling error ε is equivalent to an error in the gain α , the settling requirement for a given temperature error ΔT can be directly derived from the accuracy requirement for α , given by (5.47):

$$|\varepsilon| < \left(\frac{2}{3} \% / ^\circ\text{C} \right) \cdot \Delta T. \quad (5.49)$$

If the maximum temperature error due to incomplete settling is $0.01\ \text{K}$, for example, the maximum settling error is 0.006% .

A derivation of the settling time required to achieve a given settling error can be found in Appendix C. Applying the results presented there to the integration of ΔV_{BE} , it can be shown that the settling time required for a bipolar transistor switched between two current levels can be expressed as a multiple of a time constant τ_{max} , which is given by:

$$\tau_{max} = \frac{kT}{qI} \frac{\alpha C_S}{N_\alpha}, \quad (5.50)$$

where I is the smaller bias current of the bipolar transistor and the on-resistances of switches are assumed to be negligible compared to the impedance of the bipolar transistor. Figure C.3 in the appendix shows how many times τ_{max} is required for a given settling accuracy ε . This time is bounded by

$$t_{settle} < \tau_{max} \ln \frac{1}{\varepsilon}, \quad (5.51)$$

which is roughly $10 \cdot \tau_{max}$ for the settling accuracy of 0.006% mentioned earlier.

The time available for settling is

$$t_{settle} = \frac{1}{2N_\alpha \cdot f_{clk}}, \quad (5.52)$$

where f_{clk} is the clock frequency of the $\Sigma\Delta$ modulator. From equations (5.50)–(5.52), the minimum bias current of the bipolar transistor required for a given settling accuracy can be derived:

$$I > \frac{kT}{q} 2\alpha C_S \cdot f_{clk} \cdot \ln \frac{1}{\varepsilon} = \frac{kT}{q} 2\alpha \frac{C_S \cdot N}{t_{conv}} \cdot \ln \frac{1}{\varepsilon}. \quad (5.53)$$

For example, if $f_{clk} = 10$ kHz, $\alpha = 10$, $C_S = 10$ pF and $\varepsilon = 0.006\%$, the minimum bias current is $0.5 \mu\text{A}$. This current is independent of the number of charge transfers N_α used when integrating ΔV_{BE} : more transfers imply that a shorter time is available for settling, but also, due to the smaller sampling capacitor, that a proportionally smaller settling time is needed. In the above discussion, the finite bandwidth of the opamp has been neglected for simplicity. In practice, it should of course be taken into account.

Errors due to Finite Gain

Even if infinite time were available for settling, the sampling capacitor would not be completely discharged during phase ϕ_2 due to the finite DC gain of the opamp. This gain causes a fraction of the output voltage V_{int} to appear at the input of the opamp. As discussed in Sections 4.3.4 and 4.4.6, such leakage of the integrator's output voltage to its input limits the ENOB of the $\Sigma\Delta$ modulator. The equations presented in those sections can be used to calculate the minimum DC gain required to reduce the errors associated with leakage to negligible levels. The DC gain should also be large enough to ensure that errors and noise introduced after the first integrator are negligible when referred back to the input.

5.3.3 Noise

The noise sources of the circuit of Figure 5.10 are shown in Figure 5.11. The source $v_{n,V_{BE}}$ models the noise associated with the bipolar transistor and its current source, which is given by equation (5.13). The thermal noise of the switches equals $4kTR_{on}$ and is modelled by the sources $v_{n,sw1}$ and $v_{n,sw2}$. The noise of the opamp, finally, is modelled by $v_{n,opamp}$. As a result of these noise sources, a noise voltage v_{n,ϕ_1} appears across the sampling capacitor during phase ϕ_1 . This voltage is sampled at the end of phase ϕ_1 , resulting in a noise charge q_{n,ϕ_1} . During phase ϕ_2 , the charge on the sampling capacitor is

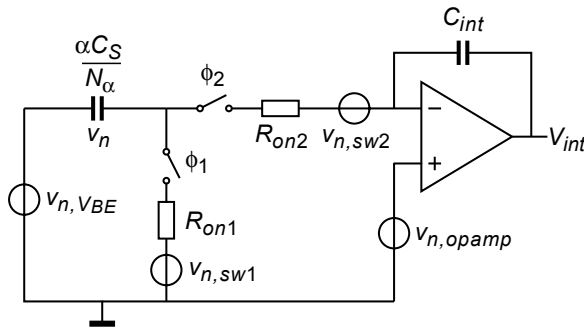


Figure 5.11. Circuit used for calculating the noise of the switched-capacitor integrator of Figure 5.10.

transferred to the integration capacitor. At the same time, a noise voltage v_{n,ϕ_2} appears across the sampling capacitor. When switch ϕ_2 opens, this voltage is sampled, resulting in a noise charge q_{n,ϕ_2} . This charge can be seen as an error in the charge transferred to the integration capacitor, so that the total noise charge $q_{n,\Delta V_{BE}}$ accumulated during the N_α charge transfers that take place within a $\Sigma\Delta$ cycle is:

$$q_{n,\Delta V_{BE}}^2 = N_\alpha \left(q_{n,\phi_1}^2 + q_{n,\phi_2}^2 \right). \quad (5.54)$$

First, q_{n,ϕ_1} will be determined. The noise voltage v_{n,ϕ_1} across the sampling capacitor at the end of phase ϕ_1 is

$$v_{n,\phi_1}^2 = v_{n,V_{BE}}^2 + v_{n,sw1}^2 = 4kT \left(\frac{1}{g_{m1}} + R_{on1} \right) B_1, \quad (5.55)$$

where g_{m1} is the transconductance of the bipolar transistor, and B_1 is the noise bandwidth during phase ϕ_1 . The latter is given by

$$B_1 = \frac{1}{4(1/g_{m1} + R_{on1}) \cdot \alpha C_S / N_\alpha}, \quad (5.56)$$

Substitution of this bandwidth in equation (5.55) gives

$$v_{n,\phi_1}^2 = N_\alpha \frac{kT}{\alpha C_S}, \quad (5.57)$$

which is the well-known ' kT/C noise'. The associated noise charge is

$$q_{n,\phi_1}^2 = \frac{kT}{N_\alpha} \alpha C_S. \quad (5.58)$$

The noise charge q_{n,ϕ_2} sampled at the end of phase ϕ_2 can be found in a similar way. The noise voltage v_{n,ϕ_2} across the sampling capacitor at the end

of phase ϕ_2 is

$$\begin{aligned} v_{n,\phi_1}^2 &= v_{n,V_{BE}}^2 + v_{n,sw2}^2 + v_{n,opamp}^2 \\ &= 4kT \left(\frac{1}{g_{m2}} + R_{on2} + \frac{\gamma}{g_{m,opamp}} \right) B_2, \end{aligned} \quad (5.59)$$

where $g_{m,opamp}$ is the transconductance the opamp's input pair. As before, the factor γ accounts for the contributions of the various signal transistors in the opamp, and therefore depends on the implementation of the opamp. As in the analysis of the CT readout (Section 5.2.3), the opamp's $1/f$ noise is neglected, because it is typically eliminated by means of dynamic offset cancellation (see Section 5.3.5). If the capacitances at the input and the output of the opamp are ignored for simplicity⁹, the noise bandwidth B_2 is given by

$$B_2 = \frac{1}{4(1/g_{m2} + R_{on2} + 1/g_{m,opamp}) \cdot \alpha C_S / N_\alpha}, \quad (5.60)$$

where $g_{m,opamp}$ is the transconductance of the opamp. The noise voltage is then

$$v_{n,\phi_2}^2 = \rho N_\alpha \frac{kT}{\alpha C_S}, \quad (5.61)$$

where the factor ρ equals

$$\rho = \frac{1/g_{m2} + R_{on2} + \gamma/g_{m,opamp}}{1/g_{m2} + R_{on2} + 1/g_{m,opamp}}. \quad (5.62)$$

The associated noise charge is

$$q_{n,\phi_2}^2 = \rho \frac{kT}{N_\alpha} \alpha C_S. \quad (5.63)$$

Substitution of q_{n,ϕ_1}^2 and q_{n,ϕ_2}^2 in equation (5.54) gives the total noise charge accumulated during a $\Sigma\Delta$ cycle:

$$q_{n,\Delta V_{BE}}^2 = (1 + \rho) \cdot kT \cdot \alpha C_S. \quad (5.64)$$

Substituting this value along with (5.44) and (5.45) in equation (5.10) yields the resulting output-referred temperature noise:

$$\begin{aligned} \sigma_{T,SC} &= A \cdot \frac{\sqrt{(1 + \rho) \cdot kT \cdot \alpha C_S}}{\alpha C_S \cdot \Delta V_{BE} + C_S \cdot V_{BE}} \sqrt{\frac{(1 - \mu)}{N}} \\ &= \frac{A}{V_{REF}} \sqrt{\frac{(1 + \rho) \cdot kT \cdot \alpha \cdot (1 - \mu)}{C_S \cdot N}}. \end{aligned} \quad (5.65)$$

⁹If, in practice, the input and output capacitances are not negligible, a more complicated expression for the noise bandwidth has to be used, but the general conclusions of the analysis will remain the same.

This shows that the output noise, to first order, is independent of the bias current levels. Higher bias currents reduce the voltage noise, but this is compensated for by an equal increase in the noise bandwidth, so that the rms noise remains the same. The output noise is also independent of the number of charge transfers N_α per $\Sigma\Delta$ cycle. It can be reduced by increasing the capacitor size C_S or the number of $\Sigma\Delta$ cycles N .

Using $A \simeq 600$ K, $V_{REF} \simeq 1.2$ V and $\mu = 0.5$, which is the bit density at $T = 300$ K, equation (5.65) can be written as

$$\sigma_{T,SC} \simeq (23 \cdot 10^{-9} \text{ K F}^{0.5}) \cdot \sqrt{\frac{(1 + \rho) \cdot \alpha}{C_S \cdot N}}. \quad (5.66)$$

Suppose, for example, that $N = 1000$, $\alpha = 10$, $\rho = 2$ (from equation (5.62) with $\gamma = 3$ and $g_{m2} = g_{m,opamp} \gg 1/R_{on2}$), and that a noise $\sigma_{T,SC} = 1$ mK is desired. The minimum required sampling capacitor size is then $C_S = 15.9$ pF.

By comparing equation (5.65) with the output-referred noise of a CT readout circuit, which is given by equation (5.33), an equivalent noise resistance $R_{n,SC}$ for the SC readout circuit can be defined such that a CT readout circuit with the same noise resistance will have the same output-referred noise. This equivalent noise resistance is

$$R_{n,SC} = \frac{1 + \rho}{2} \frac{1}{f_{clk} \cdot \alpha C_S}, \quad (5.67)$$

where f_{clk} is the clock frequency of the $\Sigma\Delta$ modulator. For the example above, the equivalent noise resistance is 940 k Ω for a clock frequency of 10 kHz. The noise performance of a SC and a CT readout circuit can now be compared using their equivalent noise resistances:

$$\frac{\sigma_{T,SC}}{\sigma_{T,CT}} = \sqrt{\frac{R_{n,SC}}{R_{n,CT}}}. \quad (5.68)$$

5.3.4 Power Consumption

The minimum power consumption of a SC readout circuit is determined by a combination of its noise and accuracy requirements. A given noise requirement, as shown in the previous section, dictates a minimum value for the product $N \cdot C_S$ to ensure that the kT/C noise is low enough. For a given conversion time t_{conv} , this product then determines the minimum bias current for the bipolar transistors and the integrator required to ensure that the settling requirements are met, as expressed by equation (5.53). This minimum bias current finally gives a lower bound on the power consumption (and hence the self-heating) of the sensor.

Suppose, using the same numbers as in the example in the previous section, that $C_S = 15.9$ pF and $N = 1000$ to obtain an output-referred noise of 1 mK. If $t_{conv} = 100$ ms, the minimum unit bias current for the bipolar transistors

can then be found from equation (5.53) as $I = 0.8 \mu\text{A}$ (using $\varepsilon = 0.006\%$). If the bias current ratio $p = 10$, the total bias current of the bipolar transistors is about $9 \mu\text{A}$.

The tight link between noise and accuracy performance and power consumption implies that there is a fundamental limitation to the accuracy that can be obtained for a given output data rate. This is because higher accuracy implies a higher average power consumption, and hence more self-heating, which in turn limits the achievable accuracy. As discussed in Section 5.1.2, the maximum supply current for self-heating below 0.01 K is in the order of a few tens of μA (depending on the package). The example above shows that the required bias currents are in the same order of magnitude, so that this fundamental accuracy limit is not reached, not even for precision sensors.

Note that the quantization noise of the $\Sigma\Delta$ modulator dictates a minimum for the number of clock cycles N , as discussed in the previous chapter. From this number, the size of the sampling capacitor can then be derived. The value of N can be increased to reduce the capacitor size (which in turn reduces the chip area), but this comes at the expense of increased errors due to charge injection (see Section 5.3.5).

As in a CT implementation (see Section 5.2.4), power can be saved by shutting down the sensor part of the time. Even if the output data rate is kept the same, this may result in a reduction of the average power consumption. If the conversion time is reduced by a factor x while the noise requirements remain the same, the clock frequency has to be increased by the same factor so as to keep the number of cycles N the same. The bias currents of the bipolar transistors and the integrator then have to be increased by the same factor to maintain the same settling behavior. Therefore, to first order, the average power consumed in the bipolar transistors and the integrator remains the same if the conversion time is reduced. However, the average power consumed in supporting circuitry such as the bias circuit does decrease, reducing the overall average power consumption.

5.3.5 Autozeroing

Autozeroed Integrator

In a SC implementation, autozeroing can be used to reduce the effects of amplifier offset and $1/f$ noise. The basic principle is simple: sample the offset of the amplifier, and then subtract it from the input signal. Figure 5.12 shows how this can be implemented in a switched-capacitor integrator [20]. As in a regular switched-capacitor integrator, the circuit operates in two phases. In a first phase ϕ_1 , the opamp is switched in unity gain. At the end of this phase (time t_1), the input voltage V_{in} is sampled on capacitor C_S with respect to the virtual ground V_x of the opamp, so that a voltage $V_{in}(t_1) - V_x(t_1)$ is stored

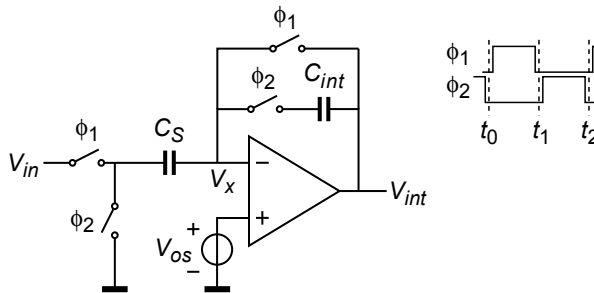


Figure 5.12. Autozeroed switched-capacitor integrator with timing of the switch signals.

on C_S . In the second phase ϕ_2 , the integration capacitor C_{int} is switched in the feedback path of the opamp, and C_S is discharged to V_x . As a result, the following charge is integrated:

$$Q = C_S \{V_{in}(t_1) + (V_x(t_2) - V_x(t_1))\}. \quad (5.69)$$

If the switches are ideal and the opamp is noiseless and has infinite open-loop gain, $V_x(t_1)$ and $V_x(t_2)$ will be equal to the offset V_{os} . The integration is then completely offset-free.

Autozeroing does not only remove the amplifier's offset but also its low-frequency noise. Due to the subtraction $V_x(t_2) - V_x(t_1)$ in (5.69), the noise is high-pass filtered, i.e. slowly varying components of the noise are filtered out along with the offset. Thus, autozeroing eliminates drift and $1/f$ noise (provided that the $1/f$ corner frequency is smaller than $f_s/2$, where f_s is the sampling frequency).

An often mentioned disadvantage of autozeroing is the aliasing of white noise due to the sampling process. This, in fact, is a problem of sampled systems in general (as is clear from the noise analysis in Section 5.3.3). In terms of white noise, the main difference between the integrator discussed in Section 5.3.3 and that of Figure 5.12 is that in the latter case, the noise of the opamp is sampled at the end of both phases. As a result, the $(1 + \rho)$ term in equation (5.65) should be replaced by a $(\rho_1 + \rho_2)$ term, where ρ_2 is defined as in equation (5.62), and ρ_1 is defined in the same way, but with g_{m2} and R_{on2} replaced by g_{m1} and R_{on1} , respectively.

Residual Offset due to Finite Gain

In practical autozeroed integrators, various non-idealities can result in residual offset. One of these non-idealities is the opamp's finite DC gain A_0 , which results in a non-zero overdrive voltage at the input of the opamp, so that V_x is

not exactly equal to V_{os} . At the end of phase ϕ_1 , V_x equals

$$V_x(t_1) = A_0 (V_{os} - V_x(t_1)) \implies V_x(t_1) = \frac{A_0}{1 + A_0} V_{os}, \quad (5.70)$$

while at the end of phase ϕ_2 , it equals

$$V_x(t_2) = V_{os} - \frac{V_{int}(t_2)}{A_0}. \quad (5.71)$$

Substitution of these expressions in (5.69) gives an integrated charge of

$$Q = C_S \left\{ V_{in}(t_1) + \frac{V_{os}}{1 + A_0} - \frac{V_{int}(t_2)}{A_0} \right\}. \quad (5.72)$$

This shows that the finite gain results in a residual offset of $V_{os}/(1 + A_0)$. The last term is a leakage error, which was already discussed in Section 5.3.2. To reduce the initial offset V_{os} , which is typically in the order of ± 1 mV, to ± 3 μ V, a DC gain of at least 50 dB is needed.

Variations of the autozeroing technique have been published that reduce errors due to finite open-loop gain. The key idea of these techniques is to sample and subtract not only the offset and noise, but also the overdrive at the input of the opamp [20]. Thus, the gain error in (5.72) can be made proportional to $1/A_0^2$, rather than to $1/A_0$. For most implementations, however, this ‘gain squaring’ is only effective for slowly changing outputs, as the overdrive sampled at the end of one clock cycle is used to compensate during the next clock cycle. In $\Sigma\Delta$ modulators, unfortunately, the integrator’s output often changes significantly from one clock cycle to the next. The usefulness of gain enhancement in this kind of applications is therefore limited, unless relatively complicated predictive gain-enhancement techniques are used [20].

Residual Offset due to Charge Injection

If the open-loop gain of the opamp is sufficiently high, the offset of an autozeroed integrator will be determined by charge injection of the switches. When a switch opens, it injects some charge into the surrounding circuitry. This charge consists of the channel charge of the switch and charge injected through overlap capacitances (also known as clock feed-through). It changes the voltages stored on C_S and C_{int} , as shown in Figure 5.13. Its magnitude depends on the size of the switches, the amplitude and slew rate of the clock signal, the size of the capacitors, and the fraction of the channel charge that flows into the capacitors [20, 27, 28]. Charge-injection errors can generally be minimized by using large capacitors and small switches (limited by other constraints, such as settling requirements). A minimum size NMOS switch ($W = 1$ μ m, $L = 0.7$ μ m) driven by 2.5 V in 0.7 μ m CMOS injects about

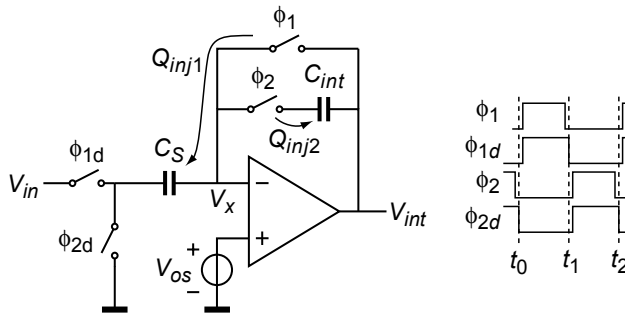


Figure 5.13. Charge injection that determines the residual offset of an autozeroed integrator.

2.5 fC. On a 10 pF capacitor, for example, this causes a voltage step of 0.25 mV, indicating that charge injection can cause considerable residual offset.

A clocking scheme with delayed falling edges, as shown in Figure 5.13, can be used to limit the charge-injection error to that of only two switches [29]. The switches driven by ϕ_1 and ϕ_2 open first, and inject some charge Q_{inj1} and Q_{inj2} on C_S and C_{int} , respectively. When the switches driven by ϕ_{1d} and ϕ_{2d} then open, the capacitors are isolated (ignoring parasitic capacitors for simplicity), so that the charge of these switches can only flow to the input source and to ground, respectively, and does not change the voltage on the capacitors¹⁰.

Using this clocking scheme, the charge injected per charge transfer is reduced to the charge injected by two switches. If multiple charge transfers are used per cycle of the $\Sigma\Delta$ modulator to implement the gain α (as shown in Figure 5.10), the total charge injected per cycle is

$$\Delta Q = 2N_\alpha Q_{inj}, \quad (5.73)$$

where N_α is the number of transfers and Q_{inj} the charge injection of a single switch. To minimize the residual offset due to charge injection, it is therefore better to implement the gain α by using a larger sampling capacitor rather than multiple charge transfers.

Even then, the residual offset is typically too large. Techniques for further reducing the offset will be discussed in Section 5.4.

5.3.6 Dynamic Element Matching

In the previous section, it was shown that it is advantageous in terms of charge injection to implement the gain α by using a larger sampling capacitor

¹⁰An additional advantage of this switching scheme is that the resulting charge-injection errors are signal independent (note that this is only true because the switch driven by ϕ_2 is placed at the opamp's input). The charge injected by switch ϕ_{1d} depends on V_{in} and thus would introduce non-linearity.

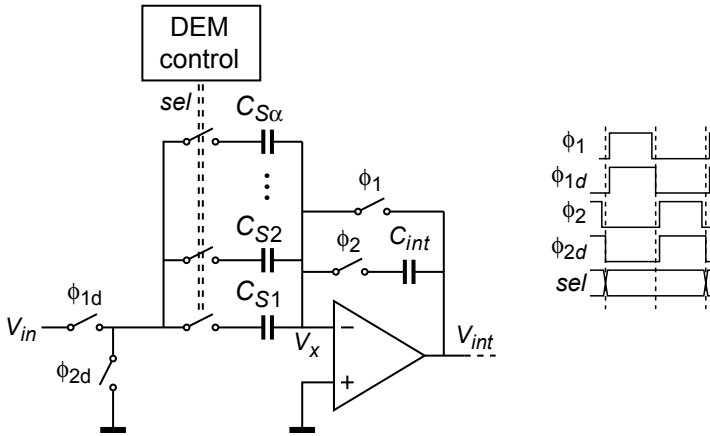


Figure 5.14. Dynamic element matching of the sampling capacitor of a SC integrator.

for ΔV_{BE} than for V_{BE} , rather than by using multiple charge transfers for ΔV_{BE} . Inaccuracy of the ratio of the sampling capacitors, however, will then lead to temperature errors. To obtain a reproducible ratio, an integer value of α should be used, so that the larger capacitor can be constructed from a parallel combination of α identical smaller capacitors. Even with a precise common-centroid layout, mismatch between these capacitors will lead to errors in the order of $\pm 0.1\%$ [15]. Using equation (5.47), this can be translated into a temperature error of $\pm 0.15^\circ\text{C}$. For precision applications, this is much too large. The error should be reduced to the $\pm 0.01^\circ\text{C}$ level, which corresponds to a maximum error in α of $\pm 0.006\%$. This implies that some form of mismatch cancellation is required.

Figure 5.14 shows how the autozeroed SC integrator of Figure 5.13 can be modified to obtain a dynamically matched sampling-capacitor ratio. A set of α sampling capacitors is used, each of which can be connected to the input using a switch. The switches are opened or closed in the non-overlap time between the clock phases ϕ_1 and ϕ_2 , so that their charge injection does not contribute to the integrated charge. If ΔV_{BE} is integrated, all capacitors are connected to the input, so that a total sampling capacitance of αC_S is created. If V_{BE} is integrated, only one of them is used. During successive cycles of the $\Sigma\Delta$ modulator in which V_{BE} is integrated, different capacitors are used, so that mismatches average out.

Assume that the values of the sampling capacitors are given by

$$C_{S_i} = C_S(1 + \delta_i), \quad 1 \leq i \leq \alpha, \quad (5.74)$$

where δ_i is the relative error of the i^{th} capacitor with respect to the average capacitance C_S , and hence

$$\sum_{i=1}^{\alpha} \delta_i = 0. \quad (5.75)$$

If the number of cycles in which V_{BE} is integrated is a multiple of α , gain errors due to mismatches completely average out. If this is not the case, an error remains that decreases with the number of cycles (see Figure 5.15). The gain error that remains after N cycles is bounded as follows:

$$\left| \frac{\alpha - \alpha_{ideal}}{\alpha} \right| = \left| \frac{1}{N} \sum_{i=1}^{N - [N/\alpha] \cdot \alpha} \delta_i \right| < \frac{\sqrt{\alpha}}{N} \delta_{max}, \quad (5.76)$$

where δ_{max} is the worst-case mismatch, i.e. all $|\delta_i| < \delta_{max}$. If, for example, $\alpha = 8$ and $\delta_{max} = 1\%$, at least 425 V_{BE} cycles are needed to reduce the error to less than $\pm 0.006\%$. Since at the lower end of the temperature range ($\mu = \frac{1}{3}$), V_{BE} is only integrated once in every 3 cycles, 1275 $\Sigma\Delta$ clock cycles per temperature conversion are needed¹¹.

The problems associated with DEM in a CT implementation (as described in 5.2.6) are much less relevant in a SC implementation. Switching transients associated with DEM are not important, as long as they have settled at the end of a clock phase when the sampling takes place. The on-resistance of switches is also much less important, because the currents flowing through switches in a SC circuit are transient currents that become negligible at the end of a clock phase.

5.4 Advanced Offset Cancellation Techniques

To process ΔV_{BE} with negligible errors, an offset in the order of a few μV is required. The residual offset of both autozeroed amplifiers and chopper amplifiers is at best of the same order of magnitude, and often larger. This section gives an overview of advanced offset cancellation techniques that be applied to achieve the additional offset reduction required to ensure that offset-related errors become negligible.

5.4.1 Charge-Injection Compensation

The residual offset of both chopper amplifiers and autozeroed amplifiers is determined by charge injection in the switches used. Charge injection can be partially compensated for by adding dummy switches that are driven by a complementary clock signal and inject an amount of charge that (partially)

¹¹This is somewhat overestimated, because the sensitivity to errors in α is smaller at the lower end of the temperature range (see Figure 3.4).

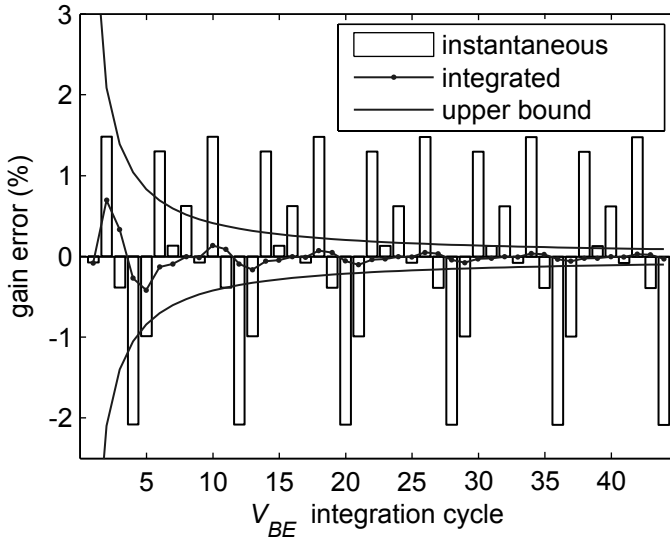


Figure 5.15. Gain error in a switched-capacitor integrator that employs DEM to average out sampling-capacitor mismatches: instantaneous gain error, gain error integrated over successive integration cycles, and upper bound for the integrated gain error given by (5.76) ($\alpha = 8$, standard deviation of the capacitor mismatch is 1%).

compensates for the charge injected by the main switch [28]. The effectiveness of such compensation depends on the matching of the injected charges. A clock signal with a high slew rate can be used to obtain a 50-50 distribution of the channel charge in the main switch [28]. A half-size dummy switch can then be used for compensation (Figure 5.16a). Since the switches are typically close to minimum size, a matching not better than 10% should be expected, so that the charge-injection-related offset is reduced by a factor 10. A disadvantage of this compensation scheme is that the main switch has to be at least twice the minimum size, and thus has a larger charge injection to start with.

An alternative way of compensating for charge injection is to use fully differential circuitry (Figure 5.16b) [27]. In that case, charge injection only results in a change in the common-mode voltage, provided that the charge injected in the two half-circuits matches. A differential voltage change only results from charge-injection *mismatch*, so that also for this compensation, a 10 times reduction of the offset due to charge injection can be expected. Advantages are that the switches now have identical dimensions (and can therefore be minimum size) and are driven by the same clock signal (no complementary clock is required). In addition to reducing charge-injection errors, incidentally, the use

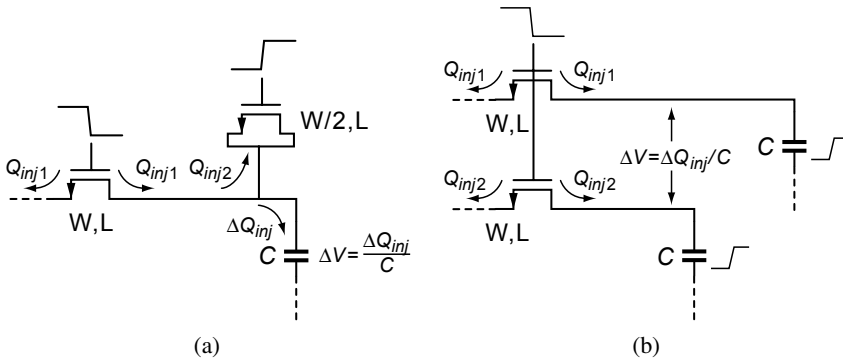


Figure 5.16. Charge-injection: (a) compensation using a half-sized dummy switch, (b) compensation using a differential circuit.

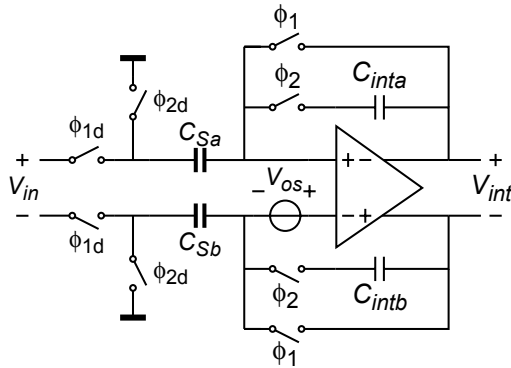


Figure 5.17. Fully differential autozeroed switched-capacitor integrator.

of fully differential circuitry has many other advantages: doubled signal swing, improved linearity, and improved common-mode and power-supply rejection [1].

Figure 5.17 shows a fully differential version of the autozeroed integrator of Figure 5.12. The offset in this implementation is determined by the charge-injection mismatch in the switches driven by ϕ_1 and ϕ_2 . For simplicity, single-ended circuits are shown in the rest of the chapter, but a fully differential implementation is to be preferred.

With fully differential circuitry, the residual offset of both chopper amplifiers and autozeroed amplifiers is at best in the order of a few μV . Additional techniques can be applied to further reduce the offset. These will be discussed in the following sections.

5.4.2 Advanced Chopping Techniques

Spike Suppression Techniques

As shown in Section 5.2.5, the residual offset of a chopper amplifier is caused by demodulated charge-injection spikes. Several techniques to suppress these spike have been published:

- *Filtering of spike harmonics.* Most of the energy of the spikes is at higher harmonics of the control signal ϕ_{ch} , while the majority of the signal energy is concentrated at the first harmonic. At the expense of a small reduction in signal gain, a significant portion of the spike energy can therefore be removed by incorporating a low-pass or band-pass filter in between the chopper switches [20, 30]. Using such techniques, residual offsets in the order of 600 nV have been obtained [31]. A disadvantage of this approach is the significant amount of extra circuitry required, such as a bandpass filter that is matched to the chopping frequency. Moreover, due to the presence of a filter in the amplifier, it is hard to apply this technique in an amplifier with overall feedback, such as the integrator of Figure 5.6.
- *Delayed demodulation.* The DC component of the demodulated spikes (and hence the residual offset) can be significantly reduced by slightly delaying the clock of the second chopper switch with respect to that of the first [32]. This effectively ‘chops’ the spikes coming from the input chopper switch into a positive and a negative part. A time delay Δt can be found such that the DC value of the chopped spikes is zero. This optimal time delay is related to the time constant τ_{spike} , which usually depends on parasitics and is hence uncontrolled. The spikes therefore have to be shaped so as to get a well-defined delay Δt . This can be done by incorporating a low-pass filter in the amplifier, while the time delay is generated using a matched low-pass filter. Offsets of the order of 1 μ V have been reported using this technique [32].
- *Chopper with guard time.* Since the spikes are usually short transients compared to the clock period, it is possible to introduce a small time gap or ‘guard time’ in the output chopper switch that prevents the spikes introduced by the input chopper switch from reaching the output. Depending on the implementation, this can be done by briefly tri-stating or shorting the output of the amplifier during the spike. Using this technique, offsets in the order of 200 nV have been obtained [16]. In a chopped V-I converter, the output current can be briefly redirected from the integrator to a ground node during the spike.

Of these techniques, the chopper with guard time is the most attractive, as it requires the least extra circuitry. An alternative is the nested-chopper technique discussed below.

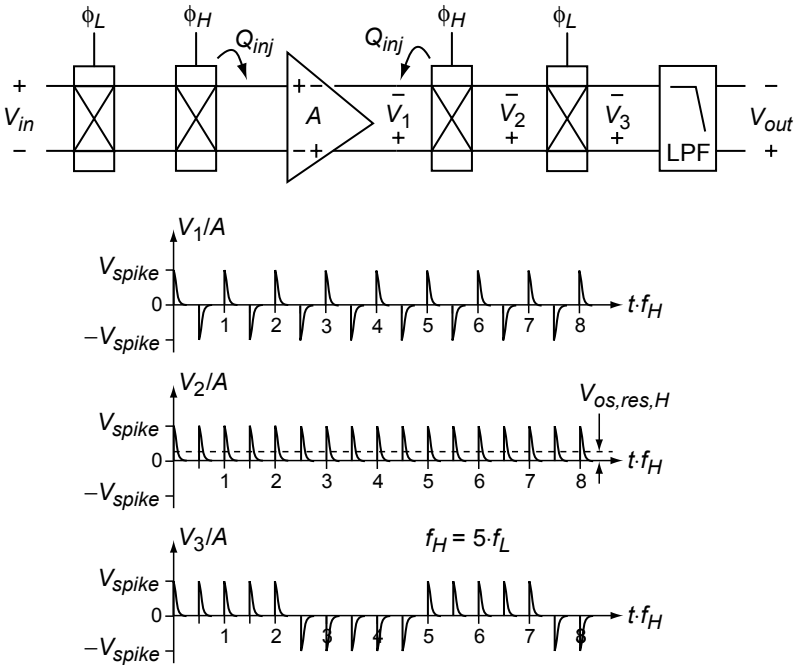


Figure 5.18. A nested-chopper amplifier: charge-injection spikes resulting from the inner chopper pair are periodically inverted by the outer chopper pair, which is clocked at a lower frequency (charge injection of the outer chopper pair is omitted for simplicity).

Nested-Chopper Technique

Since the residual offset of a chopper amplifier is proportional to the chopper frequency f_{ch} , as expressed by equation (5.41), it can be reduced by reducing f_{ch} . As mentioned in Section 5.2.5, however, f_{ch} can only be reduced below the $1/f$ noise corner frequency at the expense of increased noise at DC. The so-called *nested*-chopper technique solves this problem by using two chopper pairs, as shown in Figure 5.18 [33]. An inner chopper pair runs at a frequency f_H that is larger than the $1/f$ corner frequency. As a result, the offset and $1/f$ noise of the amplifier are modulated away from DC. A second chopper pair runs at a much lower frequency f_L and modulates the residual offset due to the inner choppers away from DC. That is, it periodically reverses the polarity of the spikes introduced by the inner choppers (see Figure 5.18). The overall residual offset is now limited by charge injection in the outer choppers, and is thus, in theory, reduced by a factor f_H/f_L .

In practice, the residual offset of a nested-chopper amplifier is not always determined by charge-injection in the outer chopper switches. If the impedances at the input of the amplifier are not matched, the positive spikes produced by the inner chopper switches will have a different time constant than the negative spikes. As a result, they will not completely average out, resulting in a residual offset that increases with f_H [33]. Another way of looking at this is that one of the offset sources (mismatch in the input impedances) lies outside the outer chopper pair and hence will not be removed by the nested chopping. When care is taken to provide matched input impedances, offsets in the order of 100 nV can be obtained using this technique [33].

The extra circuitry required to turn a chopper amplifier into a nested-chopper amplifier is very small: only an extra pair of chopper switches is needed. A disadvantage is the reduction of the usable signal bandwidth, which is now limited by f_L rather than f_H . For application in a temperature sensor, where the signal bandwidth is small anyway (in the order of 10 Hz), this is not a problem. A related disadvantage is the much lower corner frequency required in the low-pass filter, which now also has to filter out the modulated residual offset at f_L . Like the regular chopper amplifier (see Figure 5.6), however, the nested one does not require a dedicated low-pass filter, because the low-pass filtering can be performed by the $\Sigma\Delta$ ADC, as discussed in Section 4.6.

5.4.3 Advanced Autozeroing Techniques

Intermediate Offset Storage

In the autozeroed integrators discussed so far, the offset is stored on the sampling capacitors at the input of the opamp. A way to reduce the error due to charge injection is to ensure that an *amplified* version of the offset is stored. Thus, the relative error due to charge injection becomes smaller [20]. Such an amplified offset can be obtained by splitting up the opamp in two stages, and sampling the offset at an intermediate node between these stages rather than at the input [27]. The offset is then amplified by the gain of the first stage, and, in consequence, the input-referred offset due to charge injection at the intermediate node is attenuated by that same gain.

This technique is particularly useful for low-offset amplifiers or comparators. In an integrator, however, charge injection on the *integration* capacitor remains. Therefore, intermediate offset storage in an integrator results at best in a reduction of the input-referred offset by a factor of two. The same applies to offset storage on a low-sensitivity input, which is a related technique with similar advantages [34].

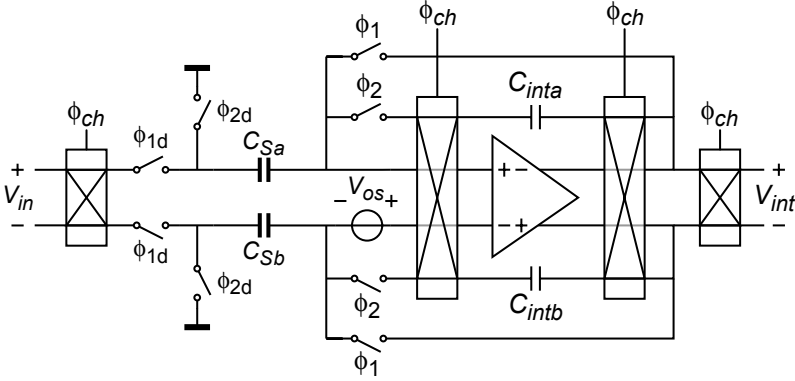


Figure 5.20. Chopped autozeroed switched-capacitor integrator.

noise floor at low frequencies (due to the chopping) and little spurious signals at harmonics of the chopping frequency (due to the autozeroing). The offset performance is then expected to be determined by charge-injection mismatch in the chopper switches and is not better than in a regular chopper amplifier.

An improvement in offset can be expected if the chopping frequency is chosen to be much lower than the autozeroing frequency. Figure 5.20 shows an implementation of an integrator based on this idea. For an amplifier, chopper switches at the input and output would suffice, but an integrator also requires that its state is inverted when its input and output are chopped [37]. In a fully differential circuit, this can easily be implemented by swapping the position of the integration capacitors.

Suppose that the residual offset of the (un-chopped) autozeroed integrator is $V_{os,res}$. The output of the integrator after N clock cycles is then

$$V_{int}(N) = N \cdot V_{os,res} + \sum_{i=1}^N V_{in}(i), \quad (5.77)$$

where $V_{in}(i)$ is the input voltage during the i^{th} clock cycle, and the integration capacitors are assumed to be initially discharged. If the integrator runs for $N_{ch}/2$ clock cycles with the chopper switches in one position, and another $N_{ch}/2$ cycles with the chopper switches in the other, the output of the integrator

after N_{ch} cycles is

$$\begin{aligned}
 V_{int}(N_{ch}) &= - \left(-V_{int} \left(\frac{N_{ch}}{2} \right) + \frac{N_{ch} \cdot V_{os,res}}{2} + \sum_{i=1+N_{ch}/2}^{N_{ch}} -V_{in}(i) \right) \\
 &= \sum_{i=1}^{N_{ch}} V_{in}(i),
 \end{aligned} \tag{5.78}$$

which shows that the offset cancels.

The residual offset of a chopped autozeroed integrator will be determined by charge injection in the chopper switches and by errors introduced during the inversion of the integrator's state. As before, these errors can be minimized by switching the chopper switches at a low frequency and in the non-overlapping time between ϕ_1 and ϕ_2 .

An advantage of the chopped autozeroed integrator of Figure 5.20 compared to the integrator of Figure 5.19, is that *any* offset sources in between the input and output choppers are eliminated. The circuit of Figure 5.19 only eliminates the dominant source, i.e. mismatch between switches ϕ_1 and ϕ_2 , but second-order sources, such as impedance mismatches and mismatches between the other switches, are not removed. In Section 7.3.6, a $\Sigma\Delta$ modulator based on chopped autozeroed integrators will be described.

5.4.4 System-Level Techniques

The offset cancellation techniques discussed so far solved the offset at the level of the integrator. An alternative is to deal with the offset at the system level. Both autozeroing and chopping can be translated to the system level.

Three-Signal Technique

Figure 5.21a shows the so-called three-signal technique, which allows for system-level compensation of offset and gain errors [38, 39]. Assume that the complete readout chain has the transfer function

$$D_{out} = K(V_s - V_{os}), \tag{5.79}$$

where D_{out} is the digital output, K the (inaccurate) system gain, V_s is the input of the readout chain, and V_{os} the (unknown) system offset. The three-signal technique consists of performing three measurements with this readout chain, corresponding to the three positions of the input switch:

$$D_{out0} = KV_{os}, \tag{5.80}$$

$$D_{out1} = K(V_{in} - V_{os}), \tag{5.81}$$

$$D_{out2} = K(V_{ref} - V_{os}), \tag{5.82}$$

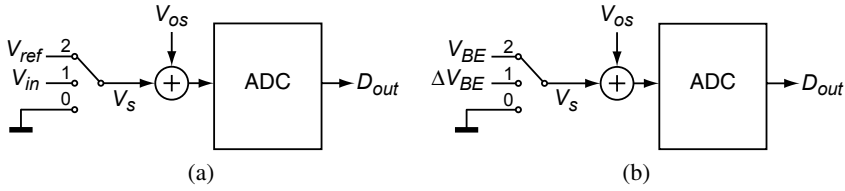


Figure 5.21. Three-signal technique: (a) principle; (b) application to temperature sensing.

where V_{ref} is an accurate reference voltage and V_{in} the measurand. The latter can then be calculated as

$$V_{in} = V_{ref} \frac{D_{out1} - D_{out0}}{D_{out2} - D_{out0}}, \quad (5.83)$$

irrespective of the values of K and V_{os} . This calculation is performed in the digital domain, for instance in a microcontroller. Thus, both offset and gain errors are eliminated.

If the system gain K can be made accurate by design (i.e. if an accurate internal reference voltage is available), the three-signal technique can be reduced to system-level autozeroing. In that case, only two measurements are needed, and the measurand V_{in} can be calculated as

$$V_{in} = (D_{out1} - D_{out0}) / K. \quad (5.84)$$

In a smart temperature sensor, the measurand and the reference voltage are combinations of V_{BE} and ΔV_{BE} (see for instance Figure 4.2a). A temperature sensor using the three-signal technique can therefore be implemented as shown in Figure 5.21b [40]: either ground, ΔV_{BE} , or V_{BE} are applied as input V_s to the readout chain, resulting in digital outputs D_{out0} , D_{out1} and D_{out2} , respectively. The desired transfer (3.6) can then be obtained from

$$\mu = \frac{\alpha \Delta V_{BE}}{V_{BE} + \alpha \Delta V_{BE}} = \frac{\alpha (D_{out1} - D_{out0})}{(D_{out2} - D_{out0}) + \alpha (D_{out1} - D_{out0})}, \quad (5.85)$$

which is again insensitive to offset and gain errors in the readout chain.

An important advantage of the three-signal technique is that *any* offset and gain errors in the whole readout chain are eliminated. Provided that the output rate is high enough, it also eliminates $1/f$ noise. The residual offset when using the three-signal technique is typically not determined by charge injection but by the resolution and linearity of the ADC. Quantization errors in the offset measurement D_{out0} limit the accuracy with which the offset is subtracted from the signal and reference measurements. Moreover, if the offset is non-linear (i.e. signal-dependent), the offset measured when shorting the input is not equal to that when a signal is applied.

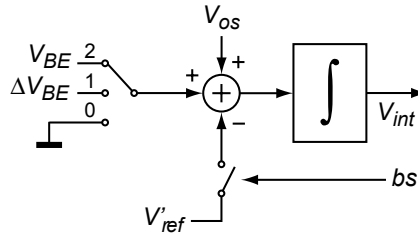


Figure 5.22. Front-end of a $\Sigma\Delta$ modulator employing the three-signal technique.

Since three AD conversions are combined to produce the final conversion result μ , the noise performance will be worse than that obtained in a system that produces μ using a single conversion. The exact noise penalty depends on the type of ADC used, and on the way the total conversion time is divided over the three conversions. An associated disadvantage is the increased dynamic range requirement for the readout chain. Since the scaling of ΔV_{BE} is performed in the digital domain, the readout circuitry in Figure 5.21b has to be able to process both an un-amplified ΔV_{BE} and V_{BE} , which effectively increases the required dynamic range by a factor α . A separate prescaler can be used [40], but this then requires its own dynamic element matching to ensure that its gain is accurate.

A final disadvantage is the added complexity in the digital domain (a division has to be performed, where the readout schemes discussed earlier only require linear scaling). Especially when the digital processing is done on-chip (as required in a smart sensor that provides a readily interpretable output), this may imply a significant amount of extra chip area in mature CMOS technology.

Figure 5.22 shows how the system of Figure 5.21b can be implemented using a $\Sigma\Delta$ modulator. The modulator uses an inaccurate local reference V'_{ref} , which determines the scale factor K mentioned above. Figure 5.22 reveals a detail of system-level autozeroing that deserves some extra attention. If the input is grounded, the only input signal applied to the integrator is the system's input-referred offset V_{os} . If this offset is negative, it cannot be measured, because only positive input voltages can be balanced by V'_{ref} . This means that either the charge-balancing scheme has to be modified to allow negative input voltages, or the offset has to be artificially increased to ensure that it is always positive.

System-Level Chopping

Figure 5.23 shows how the concept of chopping can be translated to the system level. At the input and output of the readout chain, chopper switches have been added that can reverse the signal polarity. Thus, two outputs can be generated, and the offset V_{os} can be eliminated by averaging these outputs [41].

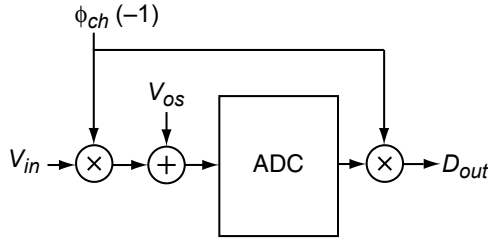


Figure 5.23. System-level chopping.

In contrast with the three-signal technique, system-level chopping requires the readout chain to have a symmetrical input range.

As for system-level autozeroing and the three-signal technique, the residual offset of system-level chopping will typically be determined by the resolution and linearity of the readout chain. A significant advantage of system-level chopping is that the input signal is converted during the full conversion time, while in the case of the mentioned alternatives, only half or one third of the conversion time is available, since the remaining time is needed for conversion of the offset and the reference voltage. Depending on the implementation, this more effective use of the conversion time may result in lower noise.

5.5 Conclusions

The performance of a smart temperature sensor that is based on a $\Sigma\Delta$ modulator is mainly determined by the first integrator of the modulator, where the charge balancing between V_{BE} and ΔV_{BE} takes place. This chapter focused on the circuit implementation of this integrator using either continuous-time (CT) or switched-capacitor (SC) techniques. These alternatives have been analyzed in terms of accuracy, noise and power consumption.

Amplifier offset is a major source of inaccuracy in both CT and SC implementations, since offsets in the μV range are required, while CMOS amplifiers have typical offsets in the mV range. Chopping and autozeroing can be used in CT and SC integrators, respectively, to reduce the offset. As a result of charge injection, however, residual offsets in the order of tens of μV remain in both cases. Advanced offset cancellation techniques, such as nested chopping, a combination of chopping and autozeroing, or system-level offset cancellation, are needed to further reduce the offset.

Mismatch is a second source of inaccuracy. Resistor or capacitor mismatch limits the initial inaccuracy of the gain α to about 0.1%, while an accuracy ten times better than that is desired. Dynamic element matching (DEM) can be used to average out mismatch-related errors. In a SC implementation, mismatch-related errors can also be avoided by using the same sampling capacitor for

both V_{BE} and ΔV_{BE} . This implies that the gain α has to be realized by using multiple charge transfers per $\Sigma\Delta$ cycle. This, however, results in larger errors due to charge injection.

CT circuits are particularly sensitive to non-linear switching transients and static errors resulting from the on-resistance of switches. Errors due to switching transients can be reduced to negligible levels either by ensuring that the transients are short compared to the clock period, or by preventing the transients from being integrated. Errors resulting from the voltage drop across switches can be prevented by using only current switches, or switches in voltage sense lines. However, this strongly limits the possible circuit configurations. SC circuits are much less sensitive to these phenomena, because they only require switching transients to have died out at the end of a clock phase. This makes it much easier to implement DEM and other dynamic error correction techniques that rely on switching in a SC implementation.

Voltage dependency of resistors and capacitors results in non-linearity of the ADC. This can be minimized using poly resistors or double-poly capacitors. As far as this non-linearity is systematic, it can be compensated for using curvature-correction techniques. Further sources of inaccuracy that affect both CT and SC implementations are leakage currents and finite amplifier gain and bandwidth.

CT implementations generally perform better in terms of noise than SC implementations. The noise bandwidth in a CT implementation is typically very small, as it is determined by the conversion time. As a result, with a conversion time of 100 ms and bias currents in the μA range, an output-referred noise in the order of only 1 mK can easily be obtained. In a SC circuit, in contrast, the full wide-band noise is sampled, which leads to kT/C noise. As a result, the noise is determined by the sampling capacitor size and the number of $\Sigma\Delta$ cycles. For 1000 cycles, tens of pF are required to obtain an output-referred noise in the order of 1 mK.

Better accuracy or noise performance usually comes at the cost of a higher power consumption. Power consumption, however, cannot be increased indefinitely, as the associated self-heating will eventually limit the accuracy. There is therefore a fundamental limit to the accuracy that can be obtained. For conversion rates in the order of 10 measurements per second, this fundamental limit does not prevent the use of either CT or SC implementations, even if noise and inaccuracy in the order of 0.01 K are required.

The power consumption in a CT implementation will typically not be determined by noise requirements, but by accuracy requirements, and will be highly implementation dependent. In a SC implementation, noise requirements dictate a minimum sampling capacitor size, and settling requirements in turn dictate a minimum bias current (and hence a minimum power consumption). In both cases, power consumption can be reduced by powering down the sensor part of the time.

In conclusion, a CT implementation is to be preferred for very low-noise or low-power sensors. A SC implementation is more attractive if dynamic error correction techniques are to be used extensively. In Chapter 7, two CT implementation and one SC implementation will be presented.

References

- [1] S. R. Norsworthy, R. Schreier, and G. C. Temes, Eds., *Delta-Sigma Data Converters: Theory, Design and Simulation*. Piscataway, New York: IEEE Press, 1997.
- [2] O. Bajdechi and J. H. Huijsing, *Systematic Design of Sigma-Delta Analog-to-Digital Converters*. Boston: Kluwer Academic Publishers, 2004.
- [3] J. V. Nicholas and D. R. White, *Traceable Temperatures*. Chichester, England: John Wiley & Sons, 1994.
- [4] G. C. M. Meijer, H. Kerkvliet, and F. N. Toth, "Non-invasive detection of micro-organisms using smart temperature sensors," *Sensors and Actuators*, vol. 18, pp. 276–281, Mar. 1994.
- [5] (2000, May) IC packages data handbook. Philips Semiconductors. [Online]. Available: <http://www.standards.philips.com/packaging/handbook/>
- [6] Thermal resistance table. Linear Technology. [Online]. Available: <http://www.linear.com/designtools/therresist.pdf>
- [7] "LM35 data sheet," National Semiconductor Corp., Nov. 2000, www.national.com.
- [8] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*. Chichester, England: John Wiley & Sons, 2001.
- [9] D. R. White and J. F. Clare, "Noise in measurements obtained by sampling," *Measurement Science and Technology*, vol. 3, no. 1, pp. 1–16, Jan. 1992.
- [10] A. J. M. Boomkamp and G. C. M. Meijer, "An accurate biomedical temperature transducer with on-chip microcomputer interfacing," in *Proc. ESSCIRC*, Sept. 1985, pp. 420–423.
- [11] G. C. M. Meijer *et al.*, "A three-terminal integrated temperature transducer with micro-computer interfacing," *Sensors and Actuators*, vol. 18, pp. 195–206, June 1989.
- [12] A. Bakker and J. H. Huijsing, "Micropower CMOS temperature sensor with digital output," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 7, pp. 933–937, July 1996.
- [13] A. Bakker and J. H. Huijsing, "A low-cost high-accuracy CMOS smart temperature sensor," in *Proc. ESSCIRC*, Sept. 1999, pp. 302–305.
- [14] A. Bakker and J. H. Huijsing, *High-Accuracy CMOS Smart Temperature Sensors*. Boston: Kluwer Academic Publishers, 2000.
- [15] A. Hastings, *The art of analog layout*. New Jersey: Prentice Hall, 2001.
- [16] Q. Huang and C. Menolfi, "A 200nV offset 6.5nV/ $\sqrt{\text{Hz}}$ noise PSD 5.6kHz chopper instrumentation amplifier in 1 μm digital CMOS," in *Dig. Techn. Papers ISSCC*, Feb. 2001, pp. 362–363, 465.

- [17] J. C. van der Meer, F. R. Riedijk, E. van Kampen, K. A. A. Makinwa, and J. H. Huijsing, "A fully integrated CMOS Hall sensor with a $3.65\mu\text{T}$ 3σ offset for compass applications," in *Dig. Techn. Papers ISSCC*, Feb. 2005, pp. 246–247.
- [18] R. Hogervorst and J. H. Huijsing, *Design of Low-Voltage, Low-Power Operational Amplifier Cells*. Boston: Kluwer Academic Publishers, 1996.
- [19] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1440, Oct. 1989.
- [20] C. C. Enz and G. C. Temes, "Circuit techniques for reducing the effects of op-amp imperfections: autozeroing, correlated double sampling, and chopper stabilization," *Proceedings of the IEEE*, vol. 84, no. 11, pp. 1584 – 1614, Nov. 1996.
- [21] K. B. Klaassen, "Digitally controlled absolute voltage division," *IEEE Transactions on Instrumentation and Measurement*, vol. 24, no. 2, pp. 106–112, June 1975.
- [22] P. C. de Jong and G. C. M. Meijer, "Absolute voltage amplification using dynamic feedback control," *IEEE Transactions on Instrumentation and Measurement*, vol. 46, no. 4, pp. 758–763, Aug. 1997.
- [23] M. Tuthill, "A switched-current, switched-capacitor temperature sensor in $0.6\text{-}\mu\text{m}$ CMOS," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 7, pp. 1117–1122, 1998.
- [24] C. Hagleitner *et al.*, "A gas detection system on a single CMOS chip comprising capacitive, calorimetric, and mass-sensitive microsensors," in *Dig. Techn. Papers ISSCC*, Feb. 2002, pp. 430–431, 479.
- [25] H. Yoshizawa, Y. Huang, P. F. Ferguson, and G. C. Temes, "MOSFET-only switched-capacitor circuits in digital CMOS technology," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 6, pp. 734–747, June 1999.
- [26] T. Tille, J. Sauerbrey, and D. S. Schmitt-Landsiedel, "A low-voltage MOSFET-only $\Sigma\Delta$ modulator for speech band applications using depletion-mode MOS-capacitors in combined series and parallel compensation," in *Proc. ISCAS*, vol. 1, May 2001, pp. 376–379.
- [27] R. C. Yen and P. R. Gray, "A MOS switched capacitor instrumentation amplifier," *IEEE Journal of Solid-State Circuits*, vol. SC-17, no. 6, pp. 1008–1013, Dec. 1982.
- [28] G. Wegmann, E. A. Vittoz, and F. Rahali, "Charge injection in analog MOS switches," *IEEE Journal of Solid-State Circuits*, vol. SC-22, no. 6, pp. 1091–1097, Dec. 1987.
- [29] D. G. Haigh and B. Singh, "A switching scheme for switched capacitor filters which reduces the effect of parasitic capacitances associated with switch control terminals," in *Proc. ISCAS*, vol. 2, May 1983, pp. 586–589.
- [30] C. Menolfi and Q. Huang, "A low-noise CMOS instrumentation amplifier for thermoelectric infrared detectors," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 7, pp. 968–976, July 1997.
- [31] C. Menolfi and Q. Huang, "A fully integrated, untrimmed CMOS instrumentation amplifier with submicrovolt offset," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 3, pp. 415–420, Mar. 1999.

- [32] C. Menolfi and Q. Huang, "A chopper modulated instrumentation amplifier with first order lowpass filter and delayed modulation scheme," in *Proc. ESSCIRC*, Sept. 1999, pp. 54–57.
- [33] A. Bakker, K. Thiele, and J. H. Huijsing, "A CMOS nested-chopper instrumentation amplifier with 100-nV offset," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 12, pp. 1877–1883, Dec. 2000.
- [34] M. Degrauwe, E. Vittoz, and I. Verbauwhede, "A micropower CMOS-instrumentation amplifier," *IEEE Journal of Solid-State Circuits*, vol. SC-20, no. 3, pp. 805–807, June 1985.
- [35] W. Lee, "A 4-channel, 18b $\Sigma\Delta$ modulator IC with chopped-offset stabilization," in *Dig. Techn. Papers ISSCC*, Feb. 1996, pp. 238–239.
- [36] A. T. K. Tang, "A $3\mu\text{V}$ -offset operational amplifier with $20\text{nV}/\sqrt{\text{Hz}}$ input noise PSD at DC employing both chopping and autozeroing," in *Dig. Techn. Papers ISSCC*, Feb. 2002, pp. 386–387.
- [37] J. Robert, G. C. Temes, V. Valencic, R. Dessoulavy, and P. Deval, "A 16-bit low-voltage CMOS A/D converter," *IEEE Journal of Solid-State Circuits*, vol. SC-22, no. 2, pp. 157–163, Apr. 1987.
- [38] M. J. S. Smith, L. Bowman, and J. D. Meindl, "Analysis, design, and performance of micropower circuits for a capacitive pressure sensor IC," *IEEE Journal of Solid-State Circuits*, vol. SC-21, no. 6, pp. 1045–1056, Dec. 1986.
- [39] G. C. M. Meijer, J. van Drecht, P. C. de Jong, and H. Neuteboom, "New concepts for smart signal processors and their applications to PSD displacement transducers," *Sensors and Actuators*, vol. 35, pp. 23–30, Oct. 1992.
- [40] S. H. Khadouri, G. C. M. Meijer, and F. M. L. van der Goes, "A CMOS interface for thermocouples with reference-junction compensation," *Analog Integrated Circuits and Signal Processing*, vol. 14, no. 3, pp. 235–248, Nov. 1997.
- [41] R. J. van der Plassche, "A sigma-delta modulator as an A/D converter," *IEEE Transactions on Circuits and Systems*, vol. 25, no. 7, pp. 510–514, July 1978.

Chapter 6

CALIBRATION TECHNIQUES

Using the design techniques presented in the previous chapters, many circuit- and device-related errors in CMOS smart temperature sensors can be sufficiently reduced. However, variations in the base-emitter voltage of the bipolar transistors (as a result of process spread and mechanical stress) will ultimately limit the achievable accuracy. Trimming, and an associated calibration procedure, are needed to correct for these variations. If high accuracy is desired, traditional calibration techniques are time-consuming and therefore costly. This chapter presents three alternative calibration techniques that combine accuracy with low production costs: batch calibration, calibration based on ΔV_{BE} measurement, and voltage reference calibration.

6.1 Introduction

In Section 3.4, various trimming techniques have been introduced to correct for temperature errors resulting from process spread of the base-emitter voltage V_{BE} . These techniques require that the error in V_{BE} is determined by means of a calibration. In this chapter, both conventional and optimized calibration techniques will be discussed. First, however, the definition of ‘calibration’ and its implications for the accuracy specification of a sensor will be reviewed, as this provides a justification for the optimized calibration techniques introduced later in the chapter.

6.1.1 Definition of Calibration

Various interpretations of the term ‘calibration’ can be found in literature. In some cases, it is used purely for the procedure of establishing the measuring error of an instrument, while in other cases it also includes adjustment of the instrument. In this work, the ISO definition of calibration will be followed [1]:

Calibration: the set of operations that establish the relationship between values indicated by a measuring instrument and the corresponding known values of a measurand.

Typically, calibration, according to this definition, results in a calibration report, with which a user can interpret the readings of the instrument, and estimate the uncertainty in these readings. In the case of a thermometer, such a report may consist of a list of temperatures along with correction factors. These correction factors indicate the deviation in the readings of the thermometer from the true temperature, along with an estimate of the uncertainty in the readings.

Calibration according to the ISO definition does not include adjustment of the instrument. Strictly speaking, adjustment of the instrument after calibration should, in fact, be avoided, as it physically alters the instrument and therefore calls for a re-calibration [1].

In a smart sensor, however, an output signal is desired that is readily interpretable. The user of a smart temperature sensor does not want to use a calibration report to apply corrections to the readings of the sensor. Therefore, a smart sensor has to be adjusted during production if the calibration shows that it does not meet its accuracy specification. This adjustment procedure will be referred to as *trimming*, which is defined as:

Trimming: the procedure of adjusting an instrument, sensor, or circuit, so as to obtain a desired output signal.

An alternative to trimming is *binning*, which means that after assembly, the sensors are sorted in different accuracy grades based on the result of a calibration. Compared to trimming, this has the advantage that no trimming hardware (non-volatile memory) is needed on the sensor chip. In its simplest form, binning comes down to disposing of sensors which do not meet the accuracy specification.

According to the above definitions, the production of a smart sensor will typically involve both a calibration step, and a trimming or binning step.

To be able to interpret the readings of an instrument in terms of standards, a calibration needs to *traceable*. Traceability is defined by ISO as follows [1]:

Traceability: the property of the result of a measurement whereby it can be related to appropriate standards, generally international or national standards, through an unbroken chain of comparisons.

The requirement of traceability implies that ‘auto-calibration’ is, according to these definitions, a contradiction in terms: a sensor can never fully calibrate itself, as there always needs to be a comparison with some other instrument in order to form a chain of comparisons that leads to a standard. In literature, the term auto-calibration is mostly used for instruments that can automatically correct for errors in *part* of the instrument by relying on the accuracy of other parts (which has often been established using an actual calibration). An example of this is the three-signal technique described in Section 5.4.4.0: it relies on

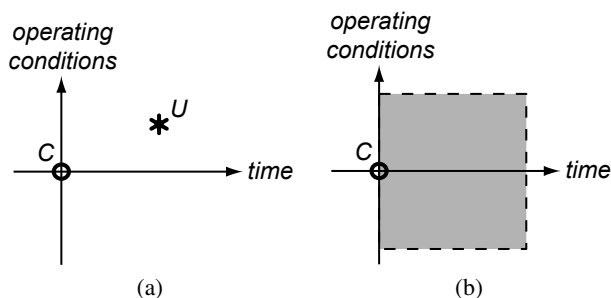


Figure 6.1. Operating space of a sensor: (a) the use U of a sensor is an extrapolation from its calibration C ; (b) accuracy specifications based on C are valid only for use within a given time from the calibration, and within a given range of operating conditions.

the availability of an accurate zero and full-scale reference in order to ‘auto-calibrate’ offset and gain errors in the rest of the system. Although such an auto-calibration can never replace an actual calibration, it can improve the stability of the instrument and thus extend the time between actual calibrations.

The traceability requirement also implies that it not useful to talk about sensors that achieve a certain accuracy *without calibration*. If a sensor has never been compared to any reference, its accuracy is unknown. It is of course possible that sensors can be produced *without trimming*. This means that the production process is so well-controlled that no adjustments are needed for the sensors to meet a given accuracy specification.

6.1.2 Extrapolation from Calibration Points

It is important to realize that any use of a sensor involves an *extrapolation* from a calibration point (Figure 6.1a). Measuring a temperature of 80°C with a sensor that was calibrated at room temperature is an example of such an extrapolation. A sensor is never used under the exact same operating conditions as during the calibration. Moreover, some time will always have passed since the calibration, during which the properties of the sensor may have changed.

Calibration data are still useful, in spite of these extrapolations, because a sensor typically has a known long-term stability, and a known sensitivity to changes in the operating conditions, which allow a manufacturer to specify the accuracy of a sensor with certain tolerances over a given period of time and over a given range of operating conditions (Figure 6.1b).

The required information about the long-term stability of a sensor and its sensitivity to changes in the operating conditions are typically determined by gathering statistical data on samples from the production process. This means that any use of a sensor to some extent relies on the stability of the production process. If, for instance, an undetected defect in the production process

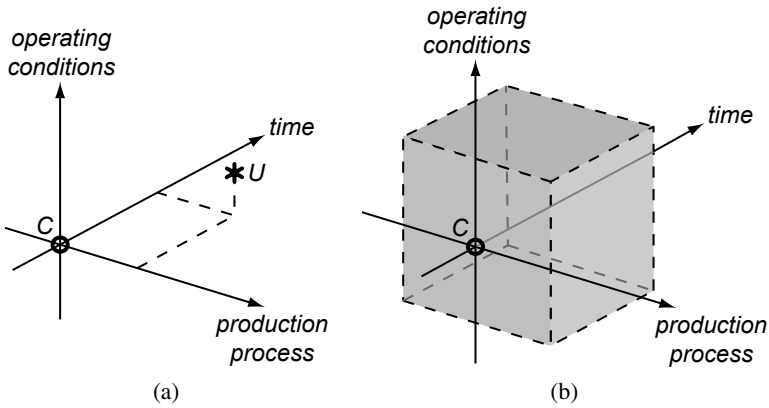


Figure 6.2. Extension of the operating space of a sensor to include variations in the production process: (a) the use U of one sensor based on the calibration C of other sensors from the same production process; (b) accuracy specifications based on C are only valid within a given time from the calibration, within a given range of operating conditions, and for sensors produced in a process not too different from the process used for making the calibrated sensors.

results in sensors with a poor long-term stability, they will not operate within specifications, no matter how good their calibration was.

An important form of extrapolation from the calibration points occurs when a sensor is calibrated and trimmed at wafer-level, and then packaged. This can be considered as a change in operating conditions, in that the mechanical stress to which the chip is exposed during calibration is different from that after packaging (see Section 2.6).

If one dimension is added to the 'operating space', so that it not only includes time and operating conditions as variables, but also the sensor's production process, we get an operating space as depicted in Figure 6.2. This extension can be interpreted as follows: provided that the production process is sufficiently stable, one cannot only make assumptions on how a sensor will perform some time after its calibration, or under different conditions than under which it was calibrated, but one can also make assumptions about the performance of *other* sensors manufactured in the same process. This is the basis for the low-cost calibration techniques presented in this chapter: from the calibration of a relatively small set of sensors from a production process with known stability, conclusions can be drawn about all the sensors from that process.

6.2 Conventional Calibration Techniques

Smart temperature sensors are usually calibrated by comparing them with a reference thermometer of known accuracy. To save production costs, this is typically done at only one temperature. It can be done either at wafer-level, or after packaging.

When calibrating at wafer-level, the temperature of a complete wafer, which may contain thousands of sensors, is stabilized and measured using a number of reference thermometers (e.g. thermistors or platinum resistors) mounted in the wafer chuck. A wafer prober then steps over the wafer, making contact to the bondpads of each of the sensor chips. It usually performs some electrical tests, takes a temperature reading from the chip, and electrically trims the sensor to adjust its reading. The time required to stabilize the temperature of the whole wafer may be significant, but it is shared by many sensors.

An important limitation of wafer-level calibration lies in the fact that the subsequent dicing and packaging can introduce temperature errors (referred to as ‘packaging shift’) [2–4]. As discussed in Section 2.6, these errors are mainly due to mechanical stress. When a chip is packaged in plastic without a stress-relieving cover layer, packaging shifts up to $\pm 0.5^\circ\text{C}$ can occur, even when relatively stress-insensitive substrate pnp transistors are used [4]. Therefore, calibration and trimming have to take place after packaging if high accuracy is to be combined with a low-cost package.

Calibration performed after packaging requires that every individual packaged sensor is brought to the same temperature as a reference thermometer. This typically means that the two are brought in good thermal contact by means of a thermally conducting medium, such as a liquid bath or a metal block. Some stabilization time will be needed, since the sensor will not precisely be at the desired temperature when it enters the calibration setup. For inaccuracies in the order of $\pm 0.1^\circ\text{C}$, this time will be much longer (tens of minutes) than the time spent on electrical tests (seconds). Unlike the case of wafer-level calibration, however, the costs associated with this long stabilization time are now associated with a single sensor, and thus dominate the total production costs.

The techniques presented in the following sections can be used to calibrate sensors after packaging without the high costs associated with an individual temperature measurement.

6.3 Batch Calibration

In an IC production process, the variations in device parameters within a production batch (intra-batch spread) are typically much smaller than the variations between batches (batch-to-batch spread). As a result, there will be a strong correlation between the temperature errors of temperature sensors from the same batch. Calibration costs can therefore be reduced by calibrating only

a limited number of samples from each batch, which are preferably evenly distributed over the wafers from the batch. The measured errors of these samples can be used to estimate the average error of the batch, which can then be used to trim all sensors from the batch [5].

The accuracy that can be obtained using such a ‘batch calibration’ depends on how accurate the estimate of the average error is, and on how large and reproducible the intra-batch spread is. Assume that N random samples are used to estimate the average error, and that the intra-batch spread has a normal distribution with standard deviation σ_{batch} . The estimate of the average error will then have a standard deviation (neglecting measurement errors) of σ_{batch}/\sqrt{N} . As a result, the standard deviation of the error of the trimmed sensors will be:

$$\sigma_{trimmed} = \sqrt{1 + \frac{1}{N}} \cdot \sigma_{batch}, \quad (6.1)$$

which means that a relatively small number of samples is enough to obtain an error spread roughly equal to the intra-batch spread. For $N = 20$, for instance, the spread only increases by 2.5%.

Batch-calibration relies on the reproducibility of the distribution of intra-batch errors. If there are outlier batches with a much larger spread, the sensors from such batches will not meet the accuracy specification. An estimate of σ_{batch} can be obtained from spread of the error of the N random samples. If this is significantly larger than expected, more samples should be tested to verify that the errors in the batch have the expected distribution.

An important question is, of course, how large an intra-batch spread σ_{batch} can be expected. Provided that the error contributions of the biasing and readout electronics have been minimized by design, this spread will be determined by the intra-batch spread of the base-emitter voltage V_{BE} of the substrate pnp transistors, and by the intra-batch spread of the bias resistor. As discussed in Sections 2.5 and 2.8.3, it is hard to draw general conclusions about the achievable intra-batch spread, as these values depend on the specific process used. In Section 7.1, a sensor will be presented that achieves a 3σ inaccuracy of $\pm 1.5^\circ\text{C}$ over the military temperature range using batch calibration. In that design, however, the readout electronics also contribute to the temperature error. An improved design will be presented in Section 7.3. This design achieves an 3σ intra-batch spread of $\pm 0.5^\circ\text{C}$ over the same operating range. If a higher accuracy is required, the sensors will have to be calibrated individually, using one of the techniques that will be introduced in the following sections.

6.4 Calibration based on ΔV_{BE} Measurement

6.4.1 Principle

As discussed in Chapter 3, the spread of the base-emitter voltage V_{BE} of substrate pnp transistors is, in principle, the only reason why CMOS smart

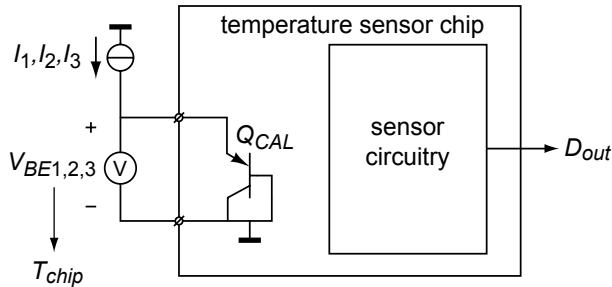


Figure 6.3. The temperature of a sensor chip T_{chip} can be derived from the base-emitter voltages of an on-chip calibration transistor Q_{CAL} .

temperature sensors need to be trimmed. An obvious idea is therefore to measure the error in V_{BE} , rather than the temperature error of the complete sensor. Unfortunately, V_{BE} is highly sensitive to temperature. Therefore, the temperature needs to be accurately measured in order to calculate the error in V_{BE} , which was exactly the problem that needed to be avoided.

A feasible alternative is to make use of the intrinsic accuracy of the *difference* in base-emitter voltage ΔV_{BE} . If generated with a well-defined current-density ratio, this difference is proportional to absolute temperature (PTAT) and, at least to first order, independent of process parameters. By measuring such a difference using an on-chip transistor and accurate *external* biasing and readout circuitry, the sensor's temperature can be determined. This temperature can then be compared with a reading of the sensor itself in order to calibrate it [6].

Since the transistor is on the same piece of silicon as the sensor circuitry, virtually no thermal settling time is required. Therefore, the calibration can be completed in the time needed to perform the voltage measurement, which is comparable to the time spent on other electrical tests (seconds). Thus, the calibration costs are kept to a minimum.

6.4.2 Implementation

Figure 6.3 shows how ΔV_{BE} -based calibration can be implemented. The sensor chip contains, in addition to the sensor circuitry, an extra substrate pnp transistor Q_{CAL} . Three externally generated bias currents $I_{1..3}$ are successively applied to the emitter of this transistor, and the resulting base-emitter voltages $V_{BE1..3}$ are measured using an accurate external voltmeter. By using three currents rather than two, errors due to series resistances (both in the transistors and in the interconnect) can be eliminated (see Section 3.7.3) [7]. From the measured base-emitter voltages, the chip's temperature T_{chip} can be calculated.

Two pins are needed in order to access Q_{CAL} from outside the chip. These pins, however, only need to be connected to Q_{CAL} during calibration. During normal operation, they can be re-used, for instance, for a digital bus interface, so that the total number of pins required on the sensor's package is not increased.

For the choice of the bias currents, the same considerations as discussed in Section 3.2 apply. The bias-current ratios should be maximized (to maximize the sensitivity), while operating the transistor in the region where its current gain is independent of the current density. Moreover, the absolute values of the bias currents should be maximized as well, so that errors due to leakage currents of ESD protection diodes are minimized. Higher bias currents also reduce the impedance of the transistor, making it less sensitive to interference. An upper bound for the bias currents is given by the error due to self-heating. The current density should be chosen such that the transistor is operated well away from its low- and high-level injection regions. For given optimal absolute values of the bias currents, the emitter area of the transistor can be adjusted to bring the current density in the desired range.

Care has to be taken to prevent errors as a result of temperature gradients on the chip. The calibration transistor should preferably be placed close to the temperature-sensitive transistors in the sensor circuitry. Also, if the temperature of the chip changes as a function of time, the calibration transistor should be either be read-out at the same time as the sensor, or some form of symmetrical time-interleaving should be used to average out differences. A simple readout sequence that cancels linear gradients is S-C-S or C-S-C, where S stands for readout of the sensor, and C for readout of the calibration transistor.

6.4.3 Accuracy

The sensitivity of the measured differences in base-emitter voltage will be in the order of $100 \mu\text{V} / ^\circ\text{C}$ (depending on the ratio of the currents used). Therefore, to obtain an absolute inaccuracy in the order of $\pm 0.1 ^\circ\text{C}$, these differences have to be measured with an inaccuracy well below $10 \mu\text{V}$. Since only differences are processed (see equation (3.90) in Section 3.7.3), offset, drift, and low-frequency noise and interference in the voltage measurement cancel. The voltage measurement, however, needs to be linear and have a resolution of more than 16 bits (given a nominal base-emitter voltage of about 600 mV). These requirements may be incompatible with production equipment. In that case, the external current sources and voltage measurement equipment may limit the achievable accuracy of the calibration.

The requirements can be relaxed if two calibration transistors are used, so that a difference ΔV_{BE} can be measured directly [8]. The required resolution is then reduced to 12 bits, and interferences will be rejected to some degree as common-mode signals. A disadvantage is that an extra pin is required, and that

the bias currents will have to be chopped between the transistors to average out mismatches.

Assuming that the differences in base-emitter voltage are accurately measured, the accuracy of the calculated temperature T_{chip} depends on the accuracy of the temperature dependency of ΔV_{BE} . In Section 7.3.9, experimental results will be presented of a temperature sensor that has been calibrated using ΔV_{BE} measurement. These results show that temperature can be calculated from a measured ΔV_{BE} with an inaccuracy of ± 0.1 °C [6]. This level of accuracy is suitable for most calibration purposes.

These results, however, rely on the reproducibility of the reverse Early voltage. As discussed in Section 2.7, the reverse Early effect causes a multiplicative error in V_{BE} and ΔV_{BE} . While this error cancels in a smart temperature sensor, due to the ratiometric nature of the measurement, it does not cancel if ΔV_{BE} is measured using external equipment. The error is typically in the order of 0.1%, or 0.3 °C [8,9]. The described calibration technique will therefore only be useful if this error is systematic. The results presented in Section 7.3.9 only cover devices from one batch. Measurements of sensors from more batches are needed to confirm that the reverse Early voltage does not spread significantly from batch to batch.

An important advantage of calibration based on ΔV_{BE} measurements, is that its accuracy is independent of that of the sensor circuitry. Therefore, a calibration transistor can easily be added to an existing design in order to reduce calibration costs with minimum design effort. The sensor's accuracy at the calibration temperature is, after trimming, purely determined by the accuracy of the calibration. Away from the calibration temperature, the sensor's accuracy will decrease, depending on the accuracy of its circuitry.

6.5 Voltage Reference Calibration

6.5.1 Principle

As discussed in the previous section, direct measurement of the error in V_{BE} without knowing the temperature is not possible, because V_{BE} is temperature dependent. Since V_{BE} is used to generate a reference voltage V_{REF} for the sensor's ADC (see Figure 3.1), calibration of this reference voltage is a feasible alternative. The calibration and trimming procedure of a temperature sensor then becomes comparable to that of a stand-alone bandgap reference: V_{REF} is measured and adjusted to its desired value. Since, ideally, the temperature coefficient of a bandgap reference is zero, knowledge of the exact calibration temperature is not required [10, 11].

Just as a ΔV_{BE} -based calibration, calibration of the voltage reference can be completed in a time comparable to that needed for other electrical tests. Thus, the extra costs for calibration are minimal.

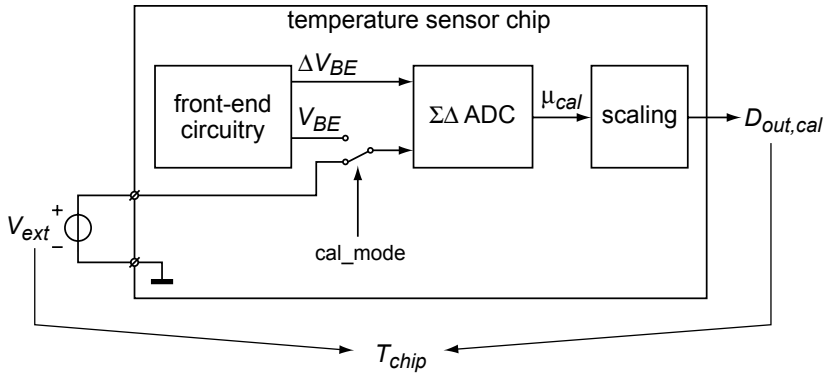


Figure 6.4. Voltage reference calibration by replacing V_{BE} with an external voltage V_{ext} .

6.5.2 Implementation

A potential problem with a direct implementation of voltage reference calibration, is that V_{REF} in a curvature-corrected temperature sensor may have a non-zero temperature coefficient (see Section 3.5.1), so that the temperature still needs to be known to determine the error in V_{REF} . Moreover, if the charge-balancing scheme of Figure 4.2b is used, the reference voltage is generated dynamically, and is not available for direct measurement.

Figure 6.4 shows how these problems can be circumvented by implementing the voltage reference calibration indirectly [12]. An external voltage V_{ext} , which can be switched in place of V_{BE} , is applied to the sensor. Thus, a temperature reading $D_{out,cal}$ is obtained. From this reading, the corresponding ratio μ_{cal} determined by the ADC can be calculated. This ratio equals:

$$\mu_{cal} = \frac{\alpha \Delta V_{BE}}{V_{ext} + \alpha \Delta V_{BE}}. \quad (6.2)$$

Since V_{ext} and α are known, ΔV_{BE} can be calculated:

$$\Delta V_{BE} = \frac{\mu_{cal}}{(1 - \mu_{cal}) \cdot \alpha} V_{ext}. \quad (6.3)$$

Thus, ΔV_{BE} has been measured indirectly using the external voltage as a reference voltage. From the value of ΔV_{BE} thus found, the sensor's temperature T_{chip} can be calculated. Finally, the sensor can be trimmed to ensure that its output D_{out} during normal operation equals the calculated temperature.

For this implementation of voltage reference calibration, one or two pins on the sensor's package are required to apply the external voltage (depending on whether the voltage is applied differentially). The pins are only needed for this purpose during the calibration procedure. During normal operation, they can

function as, for instance, digital I/O pins of the sensor's bus interface. The only extra circuitry required is a switch, which would typically be controlled via the sensor's bus interface.

6.5.3 Accuracy

The exact value of the external voltage V_{ext} is unimportant, as long as it can be determined with sufficient accuracy. To ensure that the ADC's input during calibration is roughly equal to that during normal operation, V_{ext} can be chosen roughly equal to the base-emitter voltage at the calibration temperature (typically around 550 mV). Since an error in V_{ext} directly translates to an error in the calculated temperature, V_{ext} has to be measured with a relative accuracy of $\Delta T_{max}/T_{chip}$, where ΔT_{max} is the maximum temperature error. A maximum error of $\pm 0.1^\circ\text{C}$ at a calibration temperature of 300 K, for example, implies a relative accuracy of $\pm 0.033\%$. For a value of 600 mV, this translates to an absolute accuracy of $\pm 0.2\text{ mV}$. Clearly, this is much less demanding than the absolute accuracy of $\pm 10\ \mu\text{V}$ required for the ΔV_{BE} -based calibration discussed in the previous section. Moreover, the measurement is much less sensitive to interference, because V_{ext} is generated by an external low-impedance voltage source, rather than on-chip.

The accuracy of the calculated temperature T_{chip} not only depends on the accuracy with which V_{ext} is measured, but also on how accurately the sensor implements the transfer function (6.2), and on how accurate the temperature dependency of ΔV_{BE} is. The transfer function can be made accurate using the precision techniques discussed in the previous chapters. The inaccuracy of ΔV_{BE} , as mentioned before, is expected to be $\pm 0.1^\circ\text{C}$, provided that the reverse Early voltage is reproducible between batches.

In general, the accuracy of the sensor after trimming, even at the calibration temperature, will not be better than the initial accuracy that one would expect if V_{BE} is assumed to be ideal. This means that the voltage reference calibration technique can only be applied successfully to sensors with have been designed for sufficient initial accuracy.

6.6 Conclusions

Precision CMOS smart temperature sensors have to be trimmed to correct for temperature errors resulting from spread of their bipolar transistors. Calibration is used to establish these errors. To include errors due to packaging stress, such a calibration should preferably be done after packaging. Conventional calibration techniques are then time-consuming, due to the need for thermal settling, and therefore expensive, if the errors have to be measured with an inaccuracy in the order of $\pm 0.1^\circ\text{C}$. In this chapter, three more economical calibration techniques have been presented.

The first technique is batch calibration, which is based on the fact that the spread of temperature errors within a production batch is usually smaller than the spread between batches. A limited number of samples from a batch can therefore be used to estimate the average error of that batch. This average error can then be used to trim all sensors from the batch. The residual errors after trimming will be approximately equal to the intra-batch spread. Experimental results indicate that this technique can be used to obtain an inaccuracy of about $\pm 0.5^\circ\text{C}$.

The second technique is calibration based on ΔV_{BE} measurement. It uses the fact that ΔV_{BE} is, to first order, independent of process parameters. Measured using external equipment, ΔV_{BE} of an on-chip transistor can be used to calculate the chip temperature. This voltage measurement does not take more time than other electrical tests, and therefore does not increase the production costs significantly. The extra transistor can easily be added to any existing design. The small signals involved are however sensitive to interference, and may be incompatible with production test equipment. Experimental results indicate that temperature can be measured to within $\pm 0.1^\circ\text{C}$ using ΔV_{BE} measurements, assuming that the reverse Early voltage, which affects ΔV_{BE} , does not spread significantly from batch to batch. Further experiments are needed to confirm this assumption.

The third technique is voltage reference calibration. It consists of replacing V_{BE} by an external voltage. Thus, using the on-chip ADC, the on-chip ΔV_{BE} can be measured. Again relying on the intrinsic accuracy of this ΔV_{BE} , the chip's temperature can be calculated. This technique uses a relatively large, externally applied voltage, is therefore easier to implement in a production environment. It relies, however, on the initial accuracy of the on-chip ADC, and can therefore only be applied to sensors which have been designed for high initial accuracy using the techniques discussed in the previous chapters. In that case, a similar accuracy can be obtained as with the ΔV_{BE} measurement technique.

References

- [1] J. V. Nicholas and D. R. White, *Traceable Temperatures*. Chichester, England: John Wiley & Sons, 1994.
- [2] F. Fruett and G. C. M. Meijer, *The Piezjunction Effect in Silicon Integrated Circuits and Sensors*. Boston: Kluwer Academic Publishers, May 2002.
- [3] B. Abesingha, G. A. Rincón-Mora, and D. Briggs, "Voltage shift in plastic-packaged bandgap references," *IEEE Transactions on Circuits and Systems—Part II: Analog and Digital Signal Processing*, vol. 49, no. 10, pp. 681–685, Oct. 2002.

- [4] F. Fruett, G. C. M. Meijer, and A. Bakker, "Minimization of the mechanical-stress-induced inaccuracy in bandgap voltage references," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 7, pp. 1288–1291, July 2003.
- [5] M. A. P. Pertijs, A. Bakker, and J. H. Huijsing, "A high-accuracy temperature sensor with second-order curvature correction and digital bus interface," in *Proc. ISCAS*, May 2001, pp. 368–371.
- [6] M. A. P. Pertijs and J. H. Huijsing, "Transistor temperature measurement for calibration of integrated temperature sensors," in *Proc. IMTC*, May 2002, pp. 755–758.
- [7] J. M. Audy and B. Gilbert, "Multiple sequential excitation temperature sensing method and apparatus," U.S. Patent 5 195 827, Mar. 4, 1993.
- [8] G. Wang and G. C. M. Meijer, "Temperature characteristics of bipolar transistors fabricated in CMOS technology," *Sensors and Actuators*, vol. 87, pp. 81–89, Dec. 2000.
- [9] M. A. P. Pertijs, G. C. M. Meijer, and J. H. Huijsing, "Precision temperature measurement using CMOS substrate PNP transistors," *IEEE Sensors Journal*, vol. 4, no. 3, pp. 294–300, June 2004.
- [10] G. C. M. Meijer, "Integrated circuits and components for bandgap references and temperature transducers," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, Mar. 1982.
- [11] G. A. Rincón-Mora, *Voltage References*. Piscataway, New York: IEEE Press, 2002.
- [12] M. A. P. Pertijs and J. H. Huijsing, "Digital temperature sensors and calibration thereof," U.K. Patent Application 0 507 820.9, 2005.

Chapter 7

REALIZATIONS

This chapter describes the realization of three CMOS smart temperature sensors in which the techniques introduced in the previous chapters have been applied. The first two sensors are continuous-time designs in which the most dominant errors – spread and curvature of the base-emitter voltage, and amplifier offset – have been addressed. These sensors have been implemented in a $0.7\ \mu\text{m}$ and $0.5\ \mu\text{m}$ digital CMOS process and achieve an inaccuracy of $\pm 1.5\ ^\circ\text{C}$ and $\pm 0.5\ ^\circ\text{C}$, respectively. The third sensor is a switched-capacitor design in which many more dynamic error correction techniques have been applied. This design will therefore be described in most detail. It has been implemented in a $0.7\ \mu\text{m}$ CMOS process and has an inaccuracy of $\pm 0.1\ ^\circ\text{C}$. A comparison with previous work, included at the end of the chapter, shows that this is, to date, the highest reported accuracy.

7.1 A Batch-Calibrated CMOS Smart Temperature Sensor

7.1.1 Overview

This section describes a temperature sensor that is pin-compatible with the industry-standard temperature sensor LM75 [1], the specifications of which are listed in Table 7.1. From the sensor's accuracy specifications, it is clear that not all sources of inaccuracy discussed in the previous chapters need to be addressed. Of the non-idealities discussed in Chapter 3, spread and curvature of V_{BE} are the most important. In the design of the readout circuitry, the main non-ideality that needs to be taken into account is amplifier offset. A first-order $\Sigma\Delta$ modulator meets the resolution requirement, even at modest clock frequencies.

The design presented in this section serves as an illustration of the continuous-time circuit techniques discussed in Chapter 5. It is an improved version of the

Table 7.1. Specifications of the industry-standard sensor LM75 [1].

Parameter	Value	Conditions
Inaccuracy	$\pm 2.0\text{ }^\circ\text{C}$	$-25\text{ }^\circ\text{C}$ to $100\text{ }^\circ\text{C}$
	$\pm 3.0\text{ }^\circ\text{C}$	$-55\text{ }^\circ\text{C}$ to $125\text{ }^\circ\text{C}$
Resolution	$0.5\text{ }^\circ\text{C}$	100 ms conversion time
Supply voltage	3.0 V – 5.5 V	
Supply sensitivity	$1\text{ }^\circ\text{C}/\text{V}$	
Supply current	$250\text{ }\mu\text{A}$	continuous operation

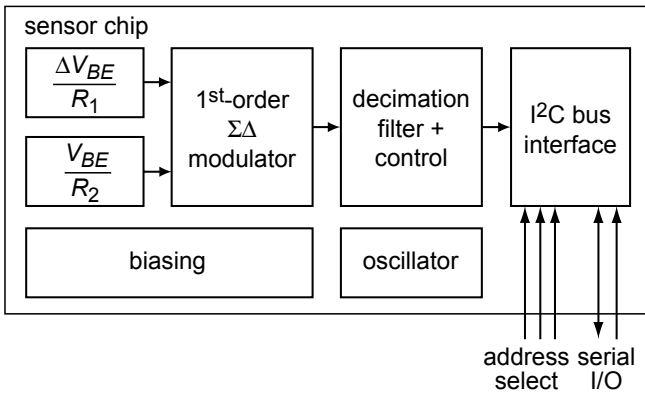


Figure 7.1. Block diagram of the improved LM75-compatible temperature sensor.

sensor of Bakker, which is described in detail in [2]. That sensor shows a systematic non-linearity of about $2\text{ }^\circ\text{C}$ over the range of $-55\text{ }^\circ\text{C}$ to $125\text{ }^\circ\text{C}$, resulting from the curvature of the bandgap reference used. In the sensor described here, the ratiometric curvature-correction technique discussed in Section 3.5.4 has been applied to correct for this systematic non-linearity.

The accuracy specification listed in Table 7.1 is at the limit of what can be obtained without trimming. From Figure 3.5, it can be seen that the $\pm 2.0\text{ }^\circ\text{C}$ requirement at $100\text{ }^\circ\text{C}$ implies that the combined batch-to-batch spread of the bias resistor and the saturation current should be less than $\pm 20\%$. While tolerances specified by foundries are usually much larger, the data presented in Section 2.5 indicate that this might just be feasible. To ensure that the sensor meets its specification without an expensive calibration per sensor, the batch calibration technique discussed in Section 6.3 will be applied.

A block diagram of the sensor is shown in Figure 7.1. Two voltage-to-current (V-I) converters generate currents proportional to ΔV_{BE} and to V_{BE} . These

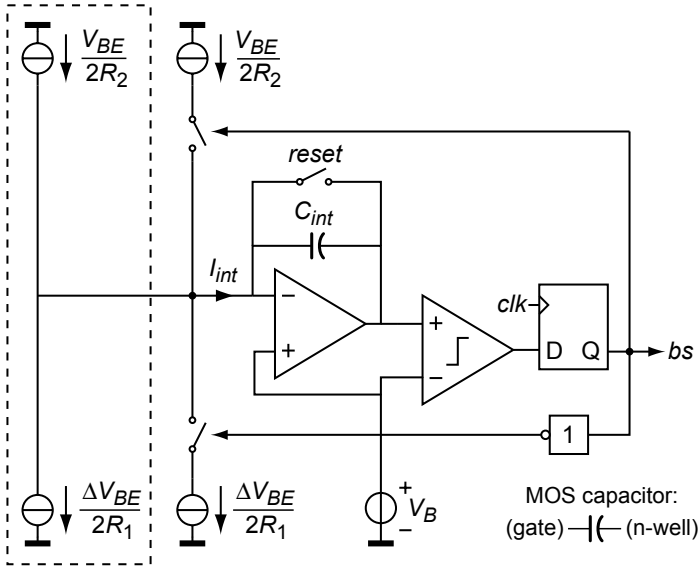


Figure 7.2. Circuit diagram of the first-order continuous-time $\Sigma\Delta$ modulator. Addition of the current sources in the dashed box results in a better use of the modulator's dynamic range.

currents are input to a first-order continuous-time $\Sigma\Delta$ modulator, the output of which is decimated using a counter to yield an 11-bit digital temperature reading. This is then communicated to the outside world using an I²C bus interface. Also on the chip are a bias circuit and an oscillator. Details of the charge-balancing operation in the modulator, the curvature correction, and the implementation of the current sources and the modulator are given in the following sections.

7.1.2 Charge-Balancing Operation

Figure 7.2 shows the circuit diagram of the first-order $\Sigma\Delta$ modulator. It consists of a continuous-time integrator, a comparator and a flip-flop. The integration capacitor C_{int} is a 60 pF MOS capacitor. Since the integrator is followed by a comparator, the non-linearity of this capacitor is not a problem. The input currents of the modulator are proportional to a base-emitter voltage V_{BE} , and proportional to a difference in base-emitter voltage ΔV_{BE} . Charge balancing is used to obtain a bitstream output of which the average value is a well-defined function of temperature.

If the current sources in the dashed box are omitted, the modulator is identical to the continuous-time implementation described in Section 5.2 (see Figure 5.3).

The average current $\overline{I_{int}}$ flowing into the integrator is then

$$\overline{I_{int}} = \mu \frac{V_{BE}}{2R_2} - (1 - \mu) \frac{\Delta V_{BE}}{2R_1}, \quad (7.1)$$

where μ is the average value of the bitstream bs . Since the feedback acts so as to null $\overline{I_{int}}$, the bitstream average μ can be written as:

$$\mu = \frac{\alpha \cdot \Delta V_{BE}}{V_{BE} + \alpha \cdot \Delta V_{BE}}, \quad (7.2)$$

where $\alpha = R_2/R_1$. This is the ratio of a PTAT voltage and a bandgap reference voltage.

As pointed out in Section 3.1.1, the ratio (7.2) results in a rather inefficient use of the modulator's dynamic range, since the extremes of the temperature range corresponds roughly to $\mu = \frac{1}{3}$ and $\mu = \frac{2}{3}$. As illustrated in Figure 3.3, only about 30% of the modulator's range is used. For a first-order $\Sigma\Delta$ modulator, this means that about 1.5 bits of resolution are lost.

This loss of resolution can be avoided by adding the current sources in the dashed box. The average input current of the integrator is then

$$\overline{I_{int}} = \frac{V_{BE}}{2R_2} - \frac{\Delta V_{BE}}{2R_1} + \mu \frac{V_{BE}}{2R_2} - (1 - \mu) \frac{\Delta V_{BE}}{2R_1}, \quad (7.3)$$

where μ is the average of the bitstream bs . Solving $\overline{I_{int}} = 0$ for μ gives

$$\mu = \frac{2\alpha \cdot \Delta V_{BE} - V_{BE}}{V_{BE} + \alpha \cdot \Delta V_{BE}}, \quad (7.4)$$

which is the more efficient ratio shown in Figure 3.3. Now, 90% of the range is used, so that the number of clock cycles needed to obtain a given resolution is reduced by a factor of three [2].

The extra currents needed for this improvement are generated by splitting the outputs of both the sinking and sourcing V-I converter in two. This comes at the cost of a possible increase of inaccuracy as a result of errors in the copied currents. For a higher-order modulator (such as the second-order modulator of the sensor discussed in the next section), the number of clock cycles is reduced by a smaller factor (e.g. only $2 \log(3) = 1.6$ for a second-order modulator). In this case, it may be more desirable to use the simpler and more accurate configuration with only two current sources.

A binary temperature reading is obtained from the bitstream using a simple counter that counts the number of ones in a sequence of N bits after the modulator has been reset. The resulting reading D_{out} is

$$D_{out} = N \cdot \mu + B. \quad (7.5)$$

The number of clock cycles N and the initial value B have been chosen so as to directly obtain a reading in degrees Celsius.

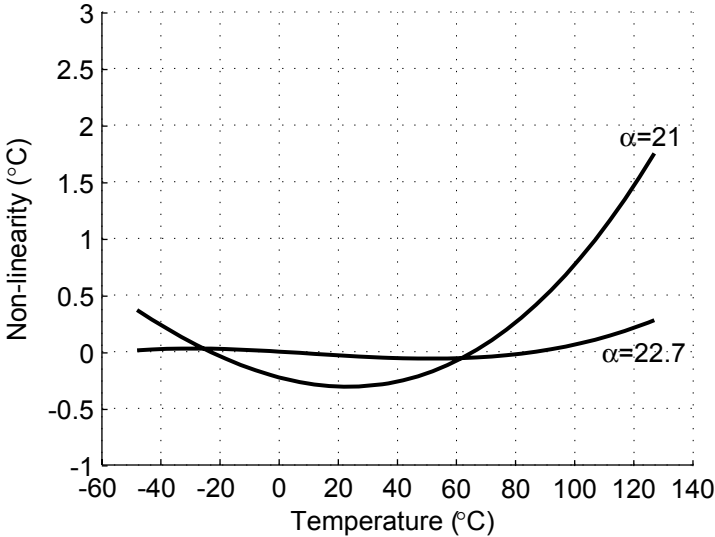


Figure 7.3. Simulated non-linearity of the sensor for a nominally temperature-independent reference ($\alpha = 21$), and a temperature-dependent reference optimized for curvature correction ($\alpha = 22.7$).

7.1.3 Curvature Correction

Figure 7.3 shows the curvature of the original design presented in [2], where a resistor ratio $\alpha = R_2/R_1 = 21$ was used. This ratio was chosen such that the denominator of equation (7.4) is a conventional bandgap reference voltage with a zero temperature coefficient at room temperature. The counter settings N and B in equation (7.5) were chosen to minimize the non-linearity around room temperature. The curvature of V_{BE} then results in a systematic non-linearity of about 2°C over the full operating temperature range.

The ratiometric curvature correction technique discussed in Section 3.5.4 has been applied to eliminate the second-order non-linearity. A slightly temperature-dependent reference has been implemented by increasing α by about 8% to 22.7. As a result, the denominator of equation (7.4) will have a positive linear temperature dependency (in addition to the curvature). This introduces a non-linearity in the bitstream average μ that cancels the non-linearity originating from the curvature of V_{BE} . What remains is a third-order non-linearity of less than 0.3°C over the full temperature range (Figure 7.3). The counter settings N and B have to be modified with respect to the original design to maintain an output in degrees Celsius.

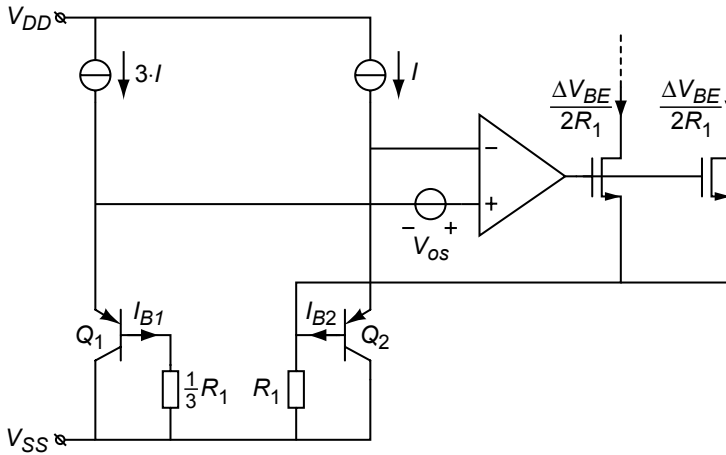


Figure 7.4. Simplified circuit diagram of the sinking V-I converter that generates a two currents proportional to ΔV_{BE} .

7.1.4 Sinking V-I Converter for ΔV_{BE}

A simplified circuit diagram of the V-I converter that implements the sinking current sources proportional to ΔV_{BE} is shown in Figure 7.4 [2]. Two substrate pnp transistors Q_1 and Q_2 are biased at a 3 : 1 current ratio. The resulting difference in base-emitter voltage ΔV_{BE} has a sensitivity of $100 \mu\text{V} / ^\circ\text{C}$. By means of a feedback loop, the emitters of the pnp transistors are kept at the same voltage, so that ΔV_{BE} is generated across a resistor $R_1 = 60 \text{ k}\Omega$ in series with the base of Q_2 . The resulting current $\Delta V_{BE} / R_1$ is split in two output currents to implement the two sinking current sources in Figure 7.2. The nominal output current at room temperature is $0.25 \mu\text{A}$.

To prevent the base current of Q_2 from affecting the output currents, a resistor $R_1/3$ is added in series with the base of Q_1 . As the base current of Q_1 is three times smaller than that of Q_2 , the base currents result in an equal voltage drop across both resistors, which is a small common-mode change that does not affect the output current.

The accuracy of the current source is mainly determined by the offset V_{os} of the opamp, which directly adds to ΔV_{BE} . Using equation (5.21), it can be shown that this offset has to be smaller than $15 \mu\text{V}$ to result in a negligible temperature error ($0.1 ^\circ\text{C}$). Since typical offsets of CMOS opamps are in the mV range, offset cancellation is required. In this design, the nested-chopper technique is used (see Section 5.4.2).

Figure 7.5 shows how a nested-chopper amplifier has been embedded in the circuit [2]. The opamp is split up into three stages, with chopper switches

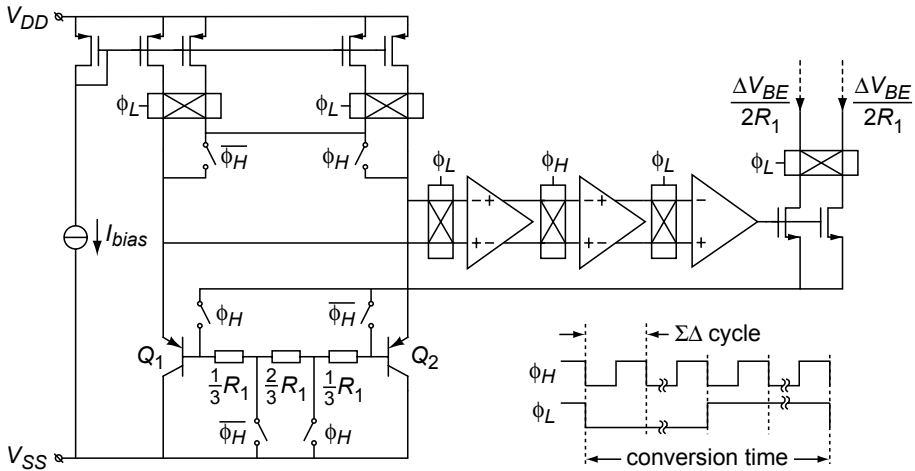


Figure 7.5. Detailed circuit diagram of the sinking V-I converter that generates two currents proportional to ΔV_{BE} .

between them. The first stage is a folded-cascode amplifier, the second stage is a differential pair, and the third stage is its current mirror load, which provides the single-ended output needed to drive the NMOS transistors.

The chopper switches are driven by two control signals: a low-frequency signal ϕ_L that runs at the conversion rate of 10 Hz, and a high-frequency signal ϕ_H that runs at the clock frequency of the $\Sigma\Delta$ modulator, 16 kHz. Offsets modulated by ϕ_H are filtered out by the continuous-time integrator of the $\Sigma\Delta$ modulator (see Section 5.2.5, Figure 5.6), while offsets modulated by ϕ_L are filtered out by the decimation filter.

The high-frequency input chopper of the nested-chopper is implemented in the current domain, by switching between a 3 : 1 and 1 : 3 current ratio. Thus, offset resulting from mismatch between the pnp transistors is also chopped. To maintain the correct feedback polarity, the connection to the output transistors is switched back and forth between the bases of Q_1 and Q_2 . As in Figure 7.4, compensation for the base currents is realized by making sure a resistor $R_1/3$ is in series with base of the transistor that carries the larger bias current. To ensure that the two output currents are equal, the two output transistors are also chopped.

The bias currents are generated by four current sources of nominally $0.5 \mu\text{A}$ each, which are dynamically matched using the same control signals ϕ_L and

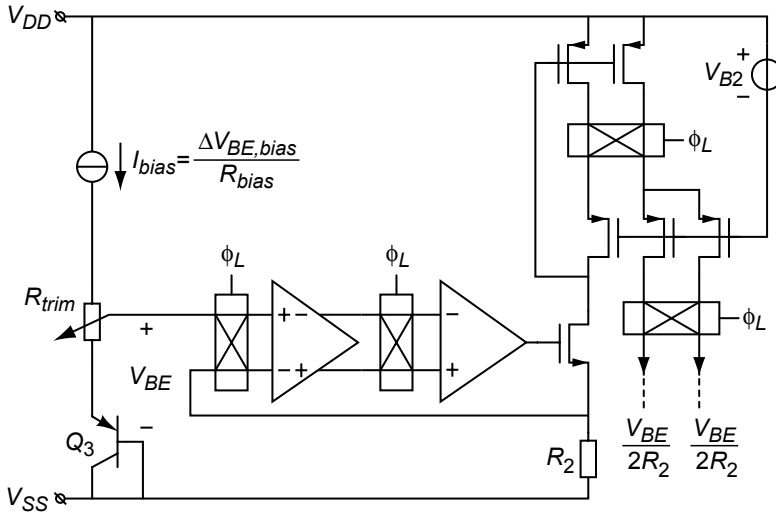


Figure 7.6. Circuit diagram of the sourcing V-I converter that generates two currents proportional to V_{BE} .

ϕ_H . Alternately, one of the current sources biases one transistor, while the remaining three bias the other. The bias currents are generated in a separate PTAT/R bias circuit (not shown).

7.1.5 Sourcing V-I Converter for V_{BE}

The V-I converter that provides the sourcing current proportional to V_{BE} is shown in Figure 7.6. It consists of a diode-connected pnp transistor with a trimmable base-emitter voltage V_{BE} , a chopped sinking V-I converter, and a chopped current mirror.

The base-emitter voltage of pnp Q_3 has been made trimmable by means of a programmable polysilicon resistor R_{trim} connected in series with its emitter (as in Figure 3.14). The bias current of nominally $1 \mu\text{A}$ is generated using a PTAT/R bias circuit based on a polysilicon bias resistor R_{bias} (the same bias circuit used to generate the bias current in Figure 7.5). The resistor R_{trim} consists of a ladder of 32 poly resistors of 900Ω each. By selecting a tap on this ladder, a programmable fraction of the PTAT voltage $\Delta V_{BE,bias}$ in the bias circuit can be added to the base-emitter voltage of Q_3 . The selection is programmable using a metal-mask change. It is wired to represent the nominal base-emitter voltage in the process, but can be modified with a single metal mask change in case the process drifts.

The resulting trimmed base-emitter voltage V_{BE} is applied to the V-I converter. Because V_{BE} has a much higher sensitivity than ΔV_{BE} , an offset below

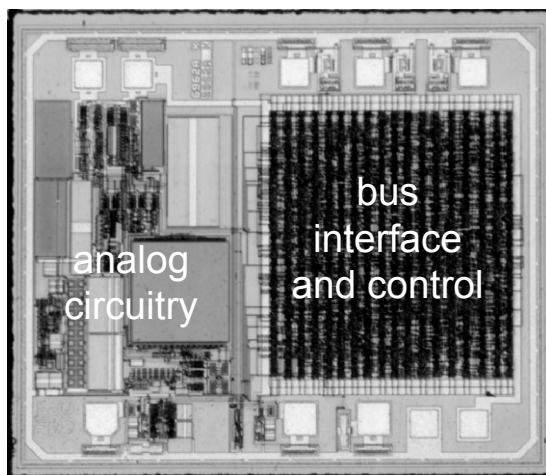


Figure 7.7. Micrograph of the temperature sensor chip.

0.3 mV is sufficient. Therefore, a regular chopper opamp driven by the low-frequency control signal ϕ_L is used. The opamp used has a folded-cascode topology; the second chopper switch is implemented in its output current mirror.

Resistor R_2 of 1360 k Ω used in the V-I converter and resistor R_1 in the sinking V-I converter for ΔV_{BE} (Figure 7.5) need to be matched, as their ratio determines the gain α . Therefore, these resistors are made from identical unit resistors of 60 k Ω , laid-out in a common-centroid pattern. The output current of the V-I converter is mirrored and split in two to implement the two sourcing current sources in Figure 7.2. The mirror transistors and cascode transistors are chopped to average out mismatches. The nominal output current of this circuit at room temperature is 0.25 μ A.

7.1.6 Experimental Results

In co-operation with Philips Semiconductors, the temperature sensor has been fabricated in a 0.7 μ m digital CMOS process. A chip micrograph is shown in Figure 7.7. The chip area is 2.8 mm², of which more than half is occupied by the I²C bus interface and control logic.

Figure 7.8 shows the measured temperature error of 50 devices from two wafers of one production batch, along with the average error and $\pm 3\sigma$ limits. The devices were mounted in plastic SO8 packages. The measurements were performed at a supply voltage of 3.3 V. The group shows a clear systematic error that can be attributed in part to batch-to-batch spread of V_{BE} . Probably it is also partly due to inaccurate modeling of the nominal base-emitter voltage

Table 7.2. Performance summary of the batch-calibrated CMOS smart temperature sensor.

Parameter	Value	Conditions
Process	0.7 μm digital CMOS	
Area	2.8 mm^2	
Inaccuracy ($\pm 3\sigma$)	$\pm 1.0^\circ\text{C}$ $\pm 1.5^\circ\text{C}$	25 $^\circ\text{C}$ –55 $^\circ\text{C}$ to 120 $^\circ\text{C}$
Resolution	0.125 $^\circ\text{C}$	100 ms conversion time
Supply voltage	2.8 V – 5.5 V	
Supply sensitivity	0.2 $^\circ\text{C}/\text{V}$	
Supply current	100 μA	continuous operation

of the pnp transistors in this process. The latter can be corrected for in later revisions by selecting a different tap on the resistor string R_{trim} by means of a metal mask change.

As a result of the relatively small intra-batch spread of about $\pm 1.0^\circ\text{C}$, the measurements can easily be brought within the desired specification by means of batch calibration (see Section 6.3). The temperature error of five random samples at 20 $^\circ\text{C}$ was used to estimate the average temperature error of the batch. Based on this estimate, all sensors were then corrected by means of the digital trimming technique discussed in Section 3.4.2. This was implemented by post-processing the readings according to equation (3.50). The resulting $\pm 3\sigma$ inaccuracy is $\pm 1.5^\circ\text{C}$ in the range of –50 $^\circ\text{C}$ to 120 $^\circ\text{C}$, which is well within the targeted LM75 specifications. The small upward curve at higher temperatures results from third-order curvature and is consistent with the simulation result of Figure 7.3. The performance of the sensor is summarized in Table 7.2.

7.2 A CMOS Smart Temperature Sensor with a 3σ Inaccuracy of $\pm 0.5^\circ\text{C}$ from –50 $^\circ\text{C}$ to 120 $^\circ\text{C}$

7.2.1 Overview

The goal of the design presented in this section is to improve both accuracy and resolution compared to the previous design. An improvement in accuracy is obtained by calibrating and trimming the sensors individually rather than per batch. So as to keep costs low, this is done using the ΔV_{BE} measurement technique described in Section 6.4: the sensor's temperature is determined from ΔV_{BE} of an extra on-chip calibration transistor, which is measured using accurate external electronics.

The resolution of the previous design is limited by the quantization noise of the first-order $\Sigma\Delta$ modulator. With a clock frequency of 16 kHz and a

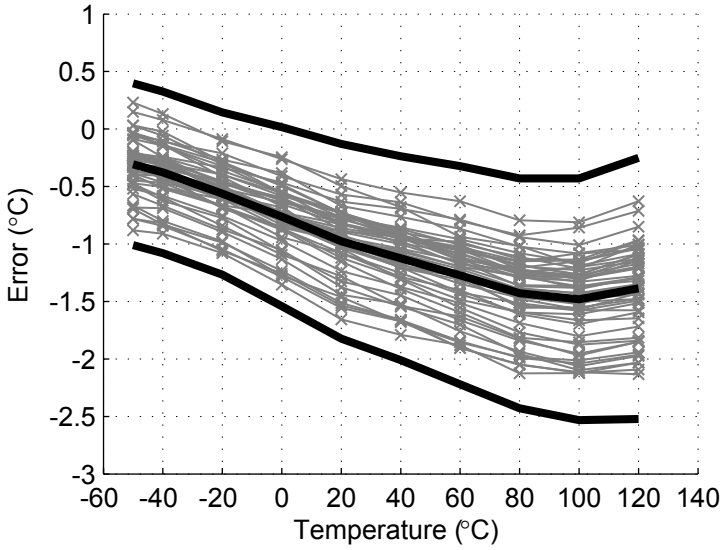


Figure 7.8. Measured temperature error before trimming, for 50 devices from 2 wafers, with average error and $\pm 3\sigma$ limits.

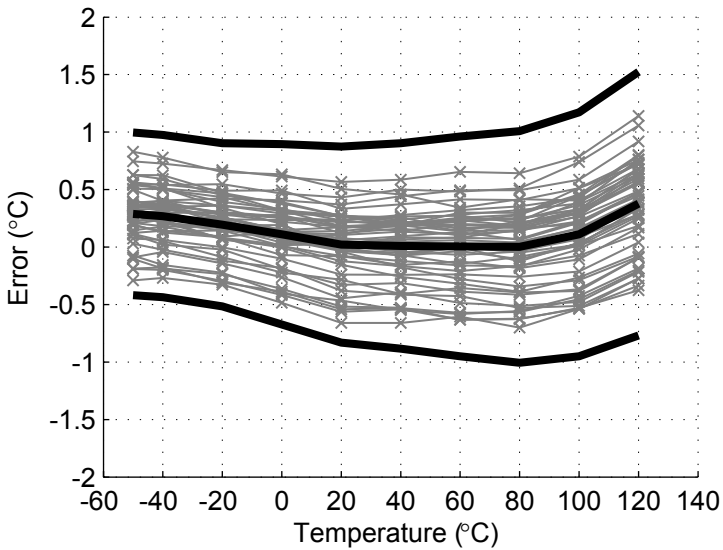


Figure 7.9. Measured temperature error of the devices in Figure 7.8 after digital trimming based on calibration of 5 random samples at 20°C .

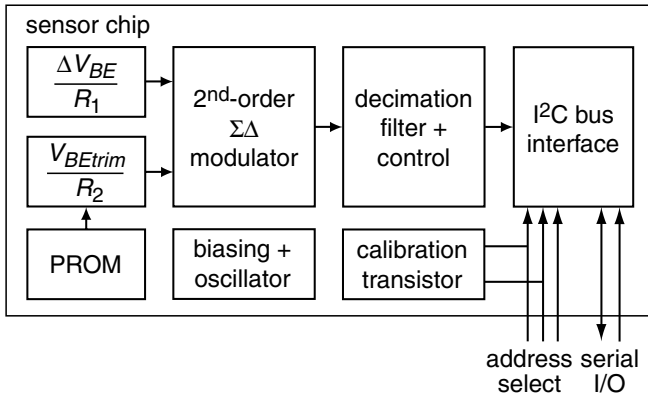


Figure 7.10. Block diagram of the $\pm 0.5^\circ\text{C}$ accurate sensor.

conversion time of 100 ms, this modulator provides a resolution of 0.125°C . In the improved design presented in this section, a resolution of 0.05°C in a conversion time of 30 ms is targeted. To obtain this resolution with a first-order modulator, the clock frequency would have to be increased to about 400 kHz. As this would lead to an undesirably high power consumption and a significant increase in errors related to charge injection, a second-order modulator is used, which can provide the desired resolution at a much lower clock frequency.

Given the higher resolution that can be obtained with a second-order modulator, there is less need to fully use its dynamic range. This means that the extra current sources in the dashed box in Figure 7.2 are not strictly necessary. Leaving them out reduces the signal-to-quantization-noise ratio by a factor three, but eliminates mismatch errors in the pairs of input currents. With the relatively low quantization noise of a second-order modulator, the threefold increase can be tolerated. The resolution requirement of 0.05°C can then still be met at a clock frequency of 16 kHz.

A block diagram of the improved sensor is shown in Figure 7.10. The V-I converters are essentially the same as in the previous design (Figures 7.5 and 7.6), except that their output currents are not split in two, and that the trimming resistor in Figure 7.6 can now be programmed via switches driven by a programmable read-only memory (PROM). Details of the $\Sigma\Delta$ modulator will be discussed in the following section.

7.2.2 Sigma-Delta Modulator

Due to the lack of linear capacitors, the loop filter of the $\Sigma\Delta$ modulator has to be implemented using MOS capacitors, whose capacitance is strongly

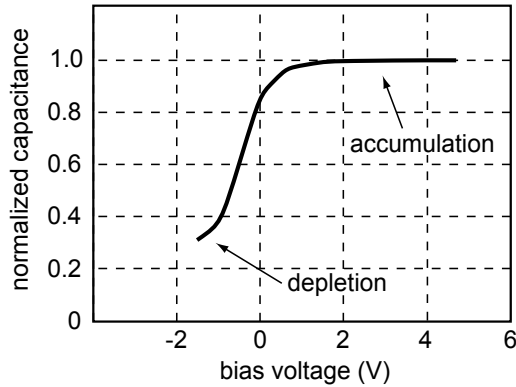


Figure 7.11. Voltage dependency of the MOS capacitors (the bias voltage is $V_{gate} - V_{well}$).

voltage-dependent (Figure 7.11). Such voltage dependency is not a problem in a continuous-time first-order modulator, because the voltage-to-charge conversion at the input is done by a resistor, and the non-linearity at the output of the integrator is unimportant, since only the sign of this output is detected by the comparator. In a second-order continuous-time modulator, however, the use of a MOS capacitor in the first integrator makes the charge-transfer from the first integrator to the second integrator highly non-linear. This is why higher-order continuous-time designs typically use poly-poly capacitors.

The use of a switched-capacitor implementation does not solve this problem. In that case, the charge transfer from the first integrator to the second integrator can be made linear by ensuring that the non-linearity of the integration capacitor of the first integrator and the sampling capacitor of the second integrator are matched. However, the voltage-to-charge conversion at the input is then no longer linear, unless special techniques are applied to linearize the sampling capacitor at the input [3].

A solution is to use a mixed continuous-time switched-capacitor architecture, as shown in Figure 7.12. The first integrator is essentially the same as that of the sensor presented in Section 7.1. The connection between the first and the second integrator is implemented using a switched-capacitor branch. The output of the first integrator is sampled by capacitor C_{S2} during clock phase ϕ_1 of a non-overlapping clock. To ensure that the charge transfer is linear, the voltage is sampled with respect to the same bias voltage V_{B1} as is used for the virtual ground of the first integrator. Moreover, capacitors C_{int1} and C_{S2} are built from identical unit capacitors so that their non-linearities are matched. During phase ϕ_2 , the charge on C_{S2} is transferred to the second integrator, which is built around capacitor C_{int2} . The non-linearity of this capacitor is not important, as the integrator is followed by a comparator that only detects the

sign of its output. The output of the comparator is clocked into a flipflop to provide the bitstream bs .

To ensure stability of the modulator and reduce the output swing of the first integrator, scaled copies of the input currents are applied to the second integrator. To prevent the integration of these copies from being disturbed by the switched-capacitor operation, the copied currents are only applied during phase ϕ_1 . The copied currents are not critical for the accuracy of the ADC, and are generated in the V-I converters using replica outputs that do not affect the accuracy of the main outputs. With this arrangement, the modulator's topology corresponds to the combined feedback-feed-forward topology discussed in Section 4.4.2, with the following loop coefficients:

$$a_1 = \frac{t_{clk}}{C_{int1}R_2}, \quad b = \frac{t_{clk}}{8C_{S2}R_2}, \quad a_2 = \frac{C_{S2}}{C_{int2}}. \quad (7.6)$$

With the chosen capacitor sizes, the ratio b/a_1 , which determines the modulator's stability, is 4. It can be concluded from Section 4.4.3 that this is a somewhat conservative choice. A higher resolution can be obtained by lowering b/a_1 to 2.

To maximize the capacitance per area of the MOS capacitors, and to avoid operating them in their most non-linear region (around 0 V, see Figure 7.11), they are biased in accumulation. This is realized using the bias voltages V_{B1} and V_{B2} . The gates of the MOS capacitors are at V_{B1} , while the feedback ensures that the average voltage on their wells is V_{B2} . Therefore, they can be biased in accumulation by choosing V_{B1} sufficiently higher than V_{B2} (in this case 1.2V).

After power-up, the modulator is brought into a well-defined state by resetting the integration capacitors. The integration capacitors could then be driven into accumulation by the feedback loop, but this might take several clock cycles (depending on the input signal). To expedite this, the integration capacitors are pre-charged using an initialization current I_{init} , as shown in Figure 7.13. This current is first applied to the first integrator, the output of which is temporarily connected to the comparator using a set of switches (not shown). After the reset, the output of the integrator is at V_{B1} . The initialization current is then applied until the outputs reaches V_{B2} , which is detected by the comparator. The initialization current is then connected to the second integrator, which is pre-charged in the same way. After that, the modulator reaches its steady-state behaviour within a few clock cycles.

7.2.3 Experimental Results

In co-operation with Philips Semiconductors, two chips have been realized in a 0.5 μm digital CMOS process: a test chip for the $\Sigma\Delta$ modulator, and a complete temperature sensor. A chip micrograph of the test chip with the $\Sigma\Delta$

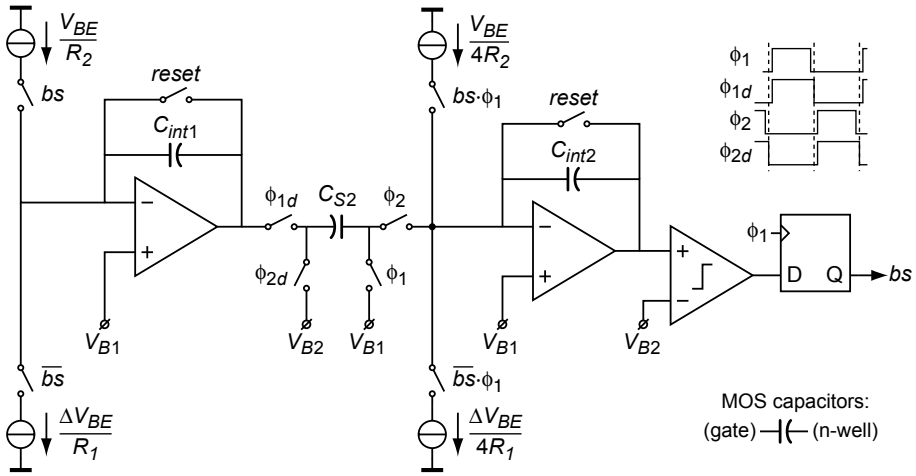


Figure 7.12. Circuit diagram of the second-order $\Sigma\Delta$ modulator; the pre-charge circuitry is omitted for clarity.

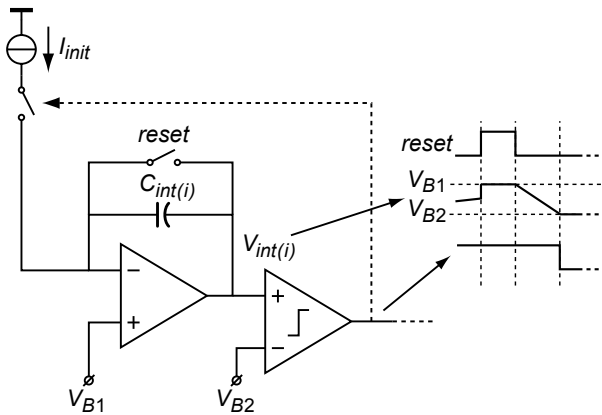


Figure 7.13. Circuit for pre-charging the integrators of the $\Sigma\Delta$ modulator.

modulator and various test circuits is shown in Figure 7.14. The chip area of the modulator is 0.23 mm^2 . The decimation filter for this test chip was realized off-chip for flexibility. A chip micrograph of the complete sensor is shown in Figure 7.15. This chip measures 2.5 mm^2 , of which about half is used for the digital bus interface and control circuitry.

Figure 7.16 shows the measured power spectrum of the bitstream of the second-order $\Sigma\Delta$ modulator. The modulator was clocked at a frequency of

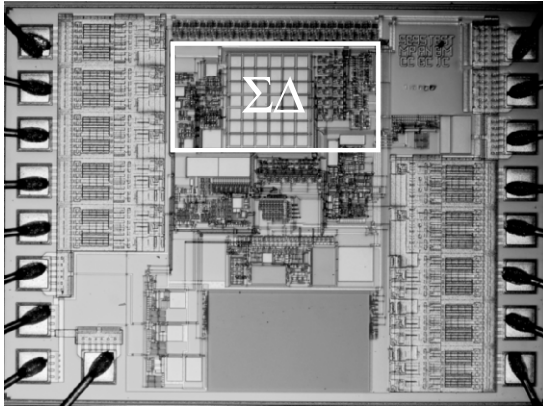


Figure 7.14. Micrograph of the testchip with the second-order $\Sigma\Delta$ modulator.

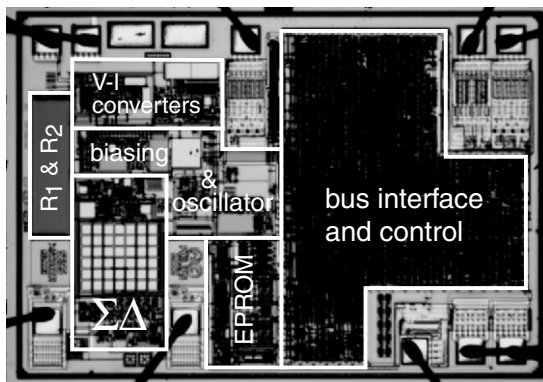


Figure 7.15. Micrograph of the temperature sensor chip.

16 kHz. The second-order noise shaping and some tones, which are the result of the DC input [4], are clearly visible.

A total of 32 samples from one processing batch were packaged in 8-pin ceramic packages. They were calibrated at room temperature by means of the ΔV_{BE} measurement technique described in Section 6.4. After trimming, the sensors were placed in an oven along with a calibrated platinum resistor in order to determine their temperature error over the full operating temperature range. During these measurements, the sensors were operated at a supply voltage of 3.3 V. The results are shown in Figure 7.17. The 3σ inaccuracy in the temperature range of -50°C to 120°C is $\pm 0.5^\circ\text{C}$. The performance of the sensors is summarized in Table 7.3.

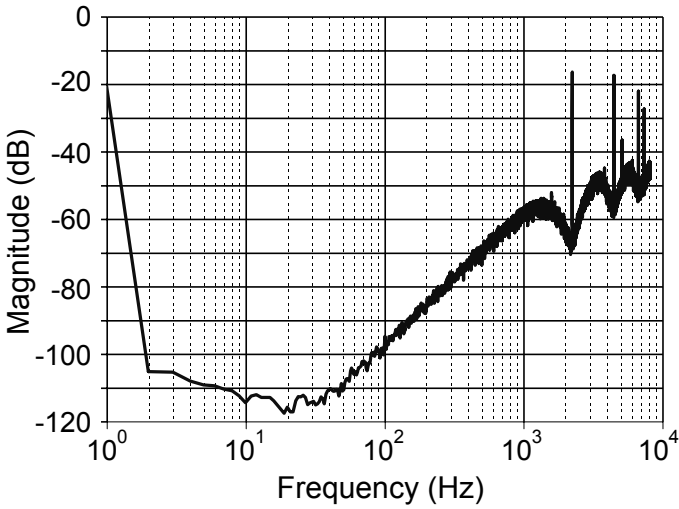


Figure 7.16. Measured power spectrum of the bitstream (16000 bits, 16x averaged, Hanning windowed)

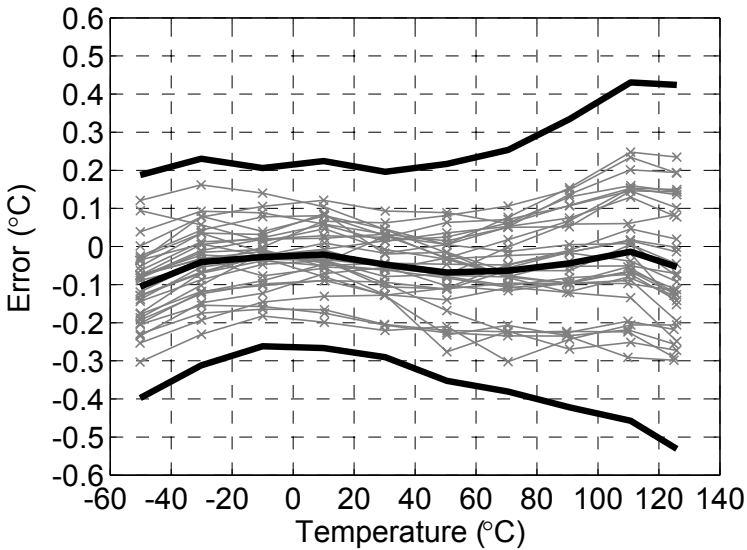


Figure 7.17. Measured temperature error of 32 samples from one batch, with average error and $\pm 3\sigma$ limits.

Table 7.3. Performance summary of the temperature sensor.

Parameter	Value	Conditions
Process	0.5 μm digital CMOS	
Area	2.5 mm^2	
Inaccuracy ($\pm 3\sigma$)	$\pm 0.3^\circ\text{C}$ $\pm 0.5^\circ\text{C}$	25 $^\circ\text{C}$ –50 $^\circ\text{C}$ to 120 $^\circ\text{C}$
Resolution	0.03 $^\circ\text{C}$	25 ms conversion time
Supply voltage	2.7 V – 5.5 V	
Supply sensitivity	0.3 $^\circ\text{C}/\text{V}$	
Supply current	130 μA	continuous operation

7.3 A CMOS Smart Temperature Sensor with a 3σ Inaccuracy of $\pm 0.1^\circ\text{C}$ from -55°C to 125°C

7.3.1 Overview

In this section, a temperature sensor design is presented that improves on the previous design in several respects. The main goal was to reduce the inaccuracy from $\pm 0.5^\circ\text{C}$ to $\pm 0.1^\circ\text{C}$ over the full military temperature range of -55°C to 125°C . In addition to that, voltage reference calibration is used rather than a calibration transistor (see Section 6.5), so as to make the calibration more suitable for implementation in a production environment, where it is much easier to apply an external voltage to a chip than to accurately measure a number of small on-chip voltages. The target specifications for the new design are summarized in Table 7.4.

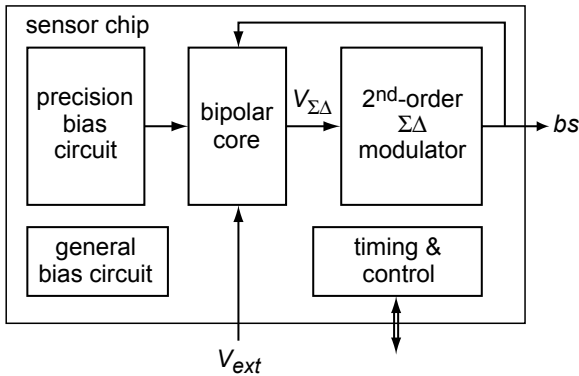
The dominant error sources in the previous design are the following:

- Inaccuracy of the gain α , which is limited by the matching of resistors R_1 and R_2 to about $\pm 0.1\%$, leads to a temperature error of about $\pm 0.15^\circ\text{C}$ (see Figure 3.4).
- Limited trimming resolution introduces an additional error of $\pm 0.1^\circ\text{C}$.
- Third-order curvature adds another 0.2°C at the high end of the temperature range, although this cannot be seen very well in Figure 7.17 in the presence of the more dominant errors.

The approach is to reduce these errors by design to a level well below $\pm 0.1^\circ\text{C}$. Mismatch errors in the gain α are reduced by means of dynamic element matching (see Section 3.2.2), the trimming resolution is increased using the modulated trimming technique (see Figure 3.17), and third-order curvature is removed using a non-linear decimation filter (see Section 4.5.4). To facilitate

Table 7.4. Target specifications

Parameter	Value	Conditions
Inaccuracy	$\pm 0.1^\circ\text{C}$	-55°C to 125°C
Resolution	0.01°C	100 ms conversion time
Supply voltage	$2.5\text{ V} - 5.5\text{ V}$	
Power consumption	$< 70\ \mu\text{W}$	1 conversion/s

Figure 7.18. Block diagram of the $\pm 0.1^\circ\text{C}$ accurate sensor.

the implementation of the dynamic element matching, a switched-capacitor rather than a continuous-time implementation is used.

A block diagram of the chip is shown in Figure 7.18. The basic topology is similar to the one presented in [5]: a single pair of substrate pnp transistors is used to generate both V_{BE} and ΔV_{BE} . This ‘bipolar core’ provides the input $V_{\Sigma\Delta}$ to a switched-capacitor second-order $\Sigma\Delta$ modulator. So as to realize the desired charge balancing, this input can be either V_{BE} or ΔV_{BE} , depending on the bitstream output bs of the modulator. The bias currents for the bipolar core are generated by a precision biasing circuit, while bias currents for all other circuits are generated by a general bias circuit. A digital block controls the timing of a temperature conversion and makes it possible to reconfigure various parts of the chip for testing purposes. Details of the charge balancing operation, the precision biasing circuit, the $\Sigma\Delta$ modulator, and the calibration procedure are discussed in the following sections.

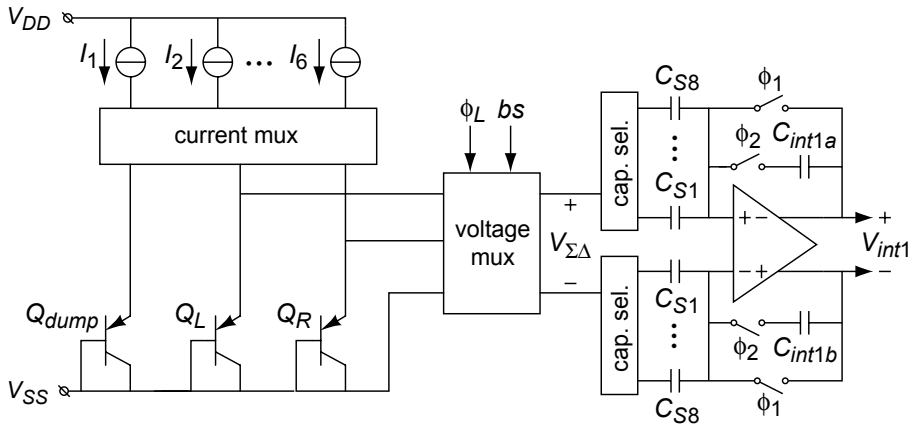


Figure 7.19. Block diagram of the temperature sensor front-end.

7.3.2 Charge-Balancing Operation

A simplified circuit diagram of the front-end of the sensor, consisting of the bipolar core and the first integrator of the $\Sigma\Delta$ modulator, is shown in Figure 7.19. It is essentially a fully differential version of the switched-capacitor implementation discussed in Section 5.3.1, built around the fully differential autozeroed integrator of Figure 5.17.

Two transistors Q_L and Q_R are used to generate V_{BE} and ΔV_{BE} . Bias currents for these transistors are provided by a set of 6 current sources, which are mirrored from the precision bias circuit, and each supply a nominal current of $1\ \mu\text{A}$. Via a current multiplexer, these currents can either be directed to the left transistor Q_L , the right transistor Q_R , or a third transistor Q_{dump} that sinks any unused currents. Thus, a 1 : 5 bias current ratio can be generated to produce ΔV_{BE} , or a programmable bias current between $0\ \mu\text{A}$ and $6\ \mu\text{A}$ to generate V_{BE} .

The sampling capacitors of the integrator have been split up in eight capacitors $C_{S1} - C_{S8}$ to be able to implement the gain α by using a smaller capacitor for sampling V_{BE} than for sampling ΔV_{BE} .

Based on the bitstream bs and control signal ϕ_L , a voltage multiplexer determines whether $+V_{BE}$, $-V_{BE}$, $+\Delta V_{BE}$, or $-\Delta V_{BE}$ is presented as input $V_{\Sigma\Delta}$ to the integrator. The timing diagram of Figure 7.20 shows how the charge-balancing scheme of Figure 4.2b has been implemented. During each cycle of the $\Sigma\Delta$ modulator, either $-V_{BE}$ or ΔV_{BE} is integrated, depending on the value of the bitstream bs :

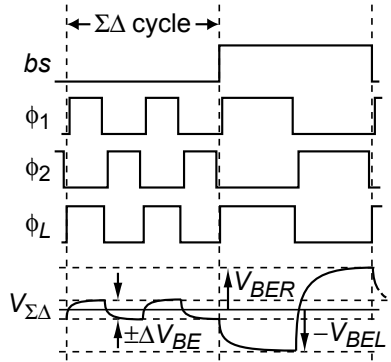


Figure 7.20. Timing of the front-end during one $\Sigma\Delta$ cycle with $bs = 0$ and one with $bs = 1$.

- If $bs = 0$, the voltage multiplexer passes ΔV_{BE} to the integrator. During phase ϕ_1 , the opamp is configured in unity gain, and $+\Delta V_{BE}$ is sampled on the parallel combination of all eight sampling capacitors $C_{S1} - C_{S8}$. During phase ϕ_2 , the integration capacitors C_{int1} are switched into the feedback path of the opamp, while the input changes to $-\Delta V_{BE}$. As a result, a charge of $2 \cdot 8 \cdot C_S \cdot \Delta V_{BE}$ is transferred to the integration capacitors, where C_S is the size of one of the sampling capacitors. Two such charge transfers are performed per cycle of the $\Sigma\Delta$ modulator, leading to a total integrated charge of

$$Q_{\Delta V_{BE}} = 32 \cdot C_S \cdot \Delta V_{BE}. \quad (7.7)$$

- If $bs = 1$, the voltage multiplexer passes $-V_{BEL}$ to the integrator during phase ϕ_1 , and V_{BER} during phase ϕ_2 , while only one of the eight sampling capacitors is connected to the input. As a result, the following charge is transferred to the integrator capacitors:

$$Q_{V_{BE}} = -C_S \cdot (V_{BEL} + V_{BER}) = -2 \cdot C_S \cdot V_{BE}, \quad (7.8)$$

where V_{BE} is the average of the base-emitter voltages of Q_L and Q_R .

The transfer of the modulator can now be derived from charge balancing: the feedback in the modulator ensures that the average integrated charge is zero:

$$(1 - \mu) \cdot (32 \cdot C_S \cdot \Delta V_{BE}) - \mu \cdot 2 \cdot C_S \cdot V_{BE} = 0, \quad (7.9)$$

where μ is the average of the bitstream. Solving for μ , we find:

$$\mu = \frac{16 \cdot \Delta V_{BE}}{V_{BE} + 16 \cdot \Delta V_{BE}} = \frac{16 \cdot \Delta V_{BE}}{V_{REF}}, \quad (7.10)$$

which is the desired ratio between a PTAT voltage and a reference voltage.

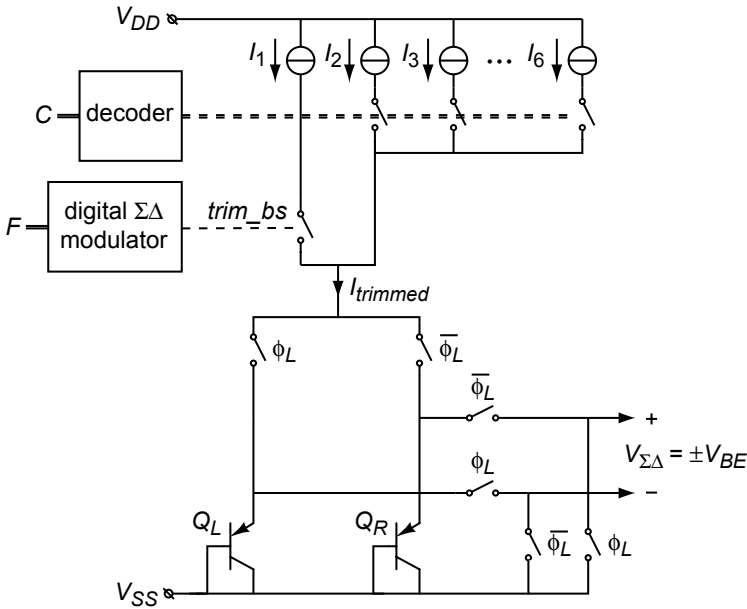


Figure 7.22. Configuration of the front-end when generating V_{BE} . The digital inputs C and F are for coarse and fine trimming of the bias current, respectively.

while the remaining currents are switched to the other transistor. The error in the resulting current ratio depends on the mismatch between the selected unit current source and the average of the other current sources. By alternating the unit current source in successive cycles of the $\Sigma\Delta$ modulator, mismatch errors are averaged out. The required averaging is performed by the integrator of the $\Sigma\Delta$ modulator [6].

7.3.4 Modulated Bias Current Trimming

The bias current used for generating V_{BE} has to be trimmed in order to compensate for the spread of the transistor's saturation current and the spread of the bias current itself. While the equivalent trimming resolution at the sensor's output has to be in the order of 0.01°C , the temperature error due to spread can be several degrees. This implies a trimming resolution of about 10 bits. Given this high trimming resolution, the modulated trimming technique introduced in Section 3.4.2 is applied [8].

Figure 7.22 shows how this has been implemented. Five of the six current sources are used for coarse trimming, and are switched on or off based on a digital input C . The sixth source is used for fine trimming, and is modulated

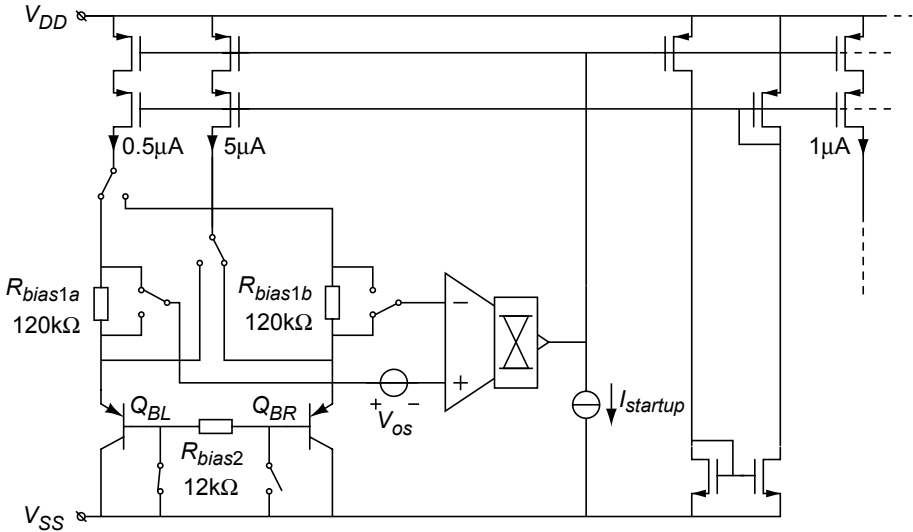


Figure 7.23. Chopped bias circuit for the bipolar core.

by the bitstream $trim_bs$ of a digital $\Sigma\Delta$ modulator. The resulting total current $I_{trimmed}$ is thus switching back and forth between C and $(C + 1)$ times the unit current of $1 \mu\text{A}$. The input F of the digital modulator can be used to program the average value of the current. The required averaging takes place in the integrator of the analog $\Sigma\Delta$ modulator.

An 8-bit first-order digital $\Sigma\Delta$ modulator is used to obtain a trimming resolution of 4 nA , which corresponds to $0.01 \text{ }^\circ\text{C}$ at the output of the sensor. A compact implementation of such a modulator is an 8-bit accumulator of which the carry bit is used to generate the bitstream [9]. The total trimming range of $0 \mu\text{A}$ – $6 \mu\text{A}$ is sufficient to compensate for practical spread of I_S and I_{bias} .

7.3.5 Precision Bias Circuit

While spread in the absolute value of the bias current can be tolerated, as it can be trimmed out, other errors in the bias current, such as its variation with the supply voltage, and spread of its temperature dependency, should be minimized. Therefore, the PTAT/R bias circuit discussed in Section 3.3.2 is used, along with the modification discussed in Section 3.6.2 needed to make the generated V_{BE} independent of the current gain of the substrate pnp transistors.

In Section 3.3.2, the accuracy of the PTAT/R bias circuit has been analyzed. From this analysis, it can be concluded that some form of offset cancellation is required in the bias circuit, because an offset below $70 \mu\text{V}$ is required to limit

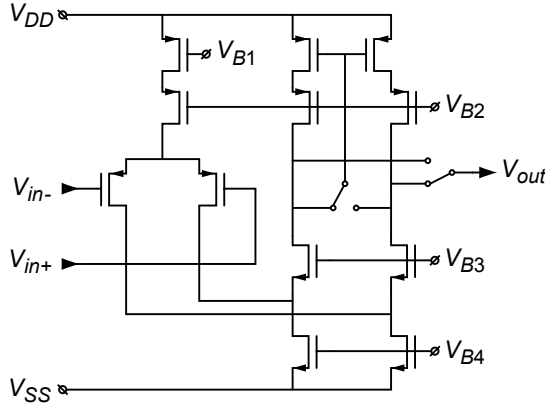


Figure 7.24. Opamp with chopped output current mirror used in the bias circuit.

the resulting temperature errors to 0.01 K. Therefore, a chopped version of the bias circuit is used, which is shown in Figure 7.23.

The input chopper switches are implemented in the current domain, biasing the transistor pair $Q_{BL} - Q_{BR}$ alternately at a 1 : 10 and a 10 : 1 current ratio. The output chopper switches have been implemented in the output current mirror of the folded-cascode opamp, as shown in Figure 7.24. The remaining switches in Figure 7.23 ensure that the resistors $R_{bias1a,b}$ and R_{bias2} are correctly connected during chopping.

As discussed in 3.6.2, the function of resistor R_{bias2} is to make the bias current dependent of the current gain of the transistors, in such a way that the generated V_{BE} is independent of this current gain.

The bias circuit is chopped synchronously with control signal ϕ_L , which determines whether the bias current is applied to Q_L or Q_R in the bipolar core (see Figure 7.22). This means that one of these transistors is biased by a current with a positive offset, while the other is biased by a current with a negative offset. As the sum of the thus generated base-emitter voltages is integrated in the $\Sigma\Delta$ modulator, the offset is almost completely eliminated:

$$\begin{aligned}
 V_{BEL} + V_{BER} &= \frac{kT}{q} \ln \left(\frac{\Delta V_{BE_bias} + V_{os}}{R_{bias1a} I_{S,L}} \right) + \frac{kT}{q} \ln \left(\frac{\Delta V_{BE_bias} - V_{os}}{R_{bias1b} I_{S,R}} \right) \\
 &\simeq 2 \frac{kT}{q} \ln \left(\frac{\Delta V_{BE_bias}}{\sqrt{R_{bias1a} \cdot R_{bias1b}} \sqrt{I_{S,L} \cdot I_{S,R}}} \right) \\
 &\quad - \frac{kT}{q} \left(\frac{V_{os}}{\Delta V_{BE_bias}} \right)^2, \tag{7.11}
 \end{aligned}$$

which is sum of the base-emitter voltage of an ‘effective transistor’ with $I_S = \sqrt{I_{S,L} \cdot I_{S,R}}$ biased using an ‘effective resistor’ $R_{bias} = \sqrt{R_{bias1a} \cdot R_{bias1b}}$, and a squared offset term. For the latter term to result in a temperature error less than 0.01 K, the initial offset V_{os} should be smaller than 2 mV, which can be achieved using careful layout. Equation (7.11) shows that there is no matching requirement between Q_L and Q_R , nor between R_{bias1a} and R_{bias1b} .

7.3.6 Sigma-Delta Modulator

Topology

A second-order $\Sigma\Delta$ modulator is used to convert the output voltages of the bipolar core into a bitstream, the average of which is a digital representation of temperature. The modulator has been implemented using two fully differential switched-capacitor integrators and a clocked comparator. A detailed circuit diagram is shown in Figure 7.25.

Ignoring the chopper switches for now, the first integrator works as described in Section 7.3.2: in a $\Sigma\Delta$ cycle in which the bitstream $bs = 0$, ΔV_{BE} is integrated using two charge transfers; if $bs = 1$, V_{BE} is integrated in one integration cycle (see the timing diagram in Figure 7.25). The sampling capacitor of the first integrator is split in 8 unit capacitors of 5 pF each. All of them are used when integrating ΔV_{BE} , while only one is used when integrating V_{BE} (selected by control signal *capsel*). Thus, a gain $\alpha = 16$ is realized. The integration capacitor C_{int1} of the first integrator is 20 pF.

At the end of a $\Sigma\Delta$ cycle, the output of the first integrator is sampled on capacitors C_F (of 2 pF), which are discharged into the second integrator at the beginning of the next $\Sigma\Delta$ cycle. The integration capacitor of the second integrator is 3 pF. To ensure stability of the modulator, a feed-forward branch from the input to the second integrator is used. The sampling capacitors C_B (of 1 pF) and $7 \cdot C_B$ of this branch are switched with the same timing as the input sampling capacitors. With this arrangement, the modulator’s topology corresponds to the combined feedback–feed-forward topology discussed in Section 4.4.2, with the following loop coefficients:

$$a_1 = \frac{C_S}{C_{int1}} = \frac{1}{4}, \quad b = \frac{C_B}{C_F} = \frac{1}{2}, \quad a_2 = \frac{C_F}{C_{int2}} = \frac{2}{3}. \quad (7.12)$$

With the chosen capacitor sizes, the ratio b/a_1 , which determines the modulator’s stability, is 2, which is the optimal value found in Section 4.4.3.

Implementation Details

The first integrator has been implemented around a fully differential folded-cascode opamp (Figure 7.26). The settling behaviour and DC gain of this opamp are important for the overall performance of the modulator. The circuit

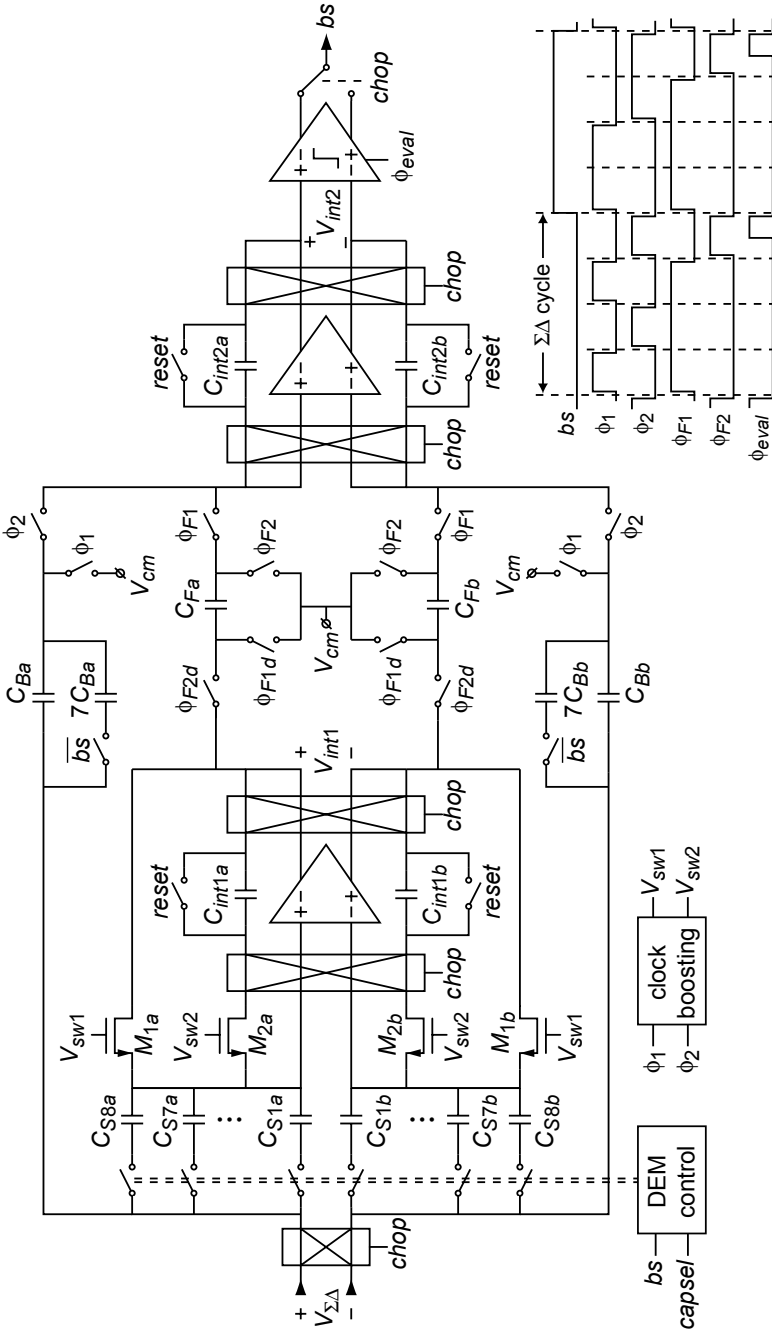


Figure 7.25. Circuit diagram of the chopped second-order $\Sigma\Delta$ modulator

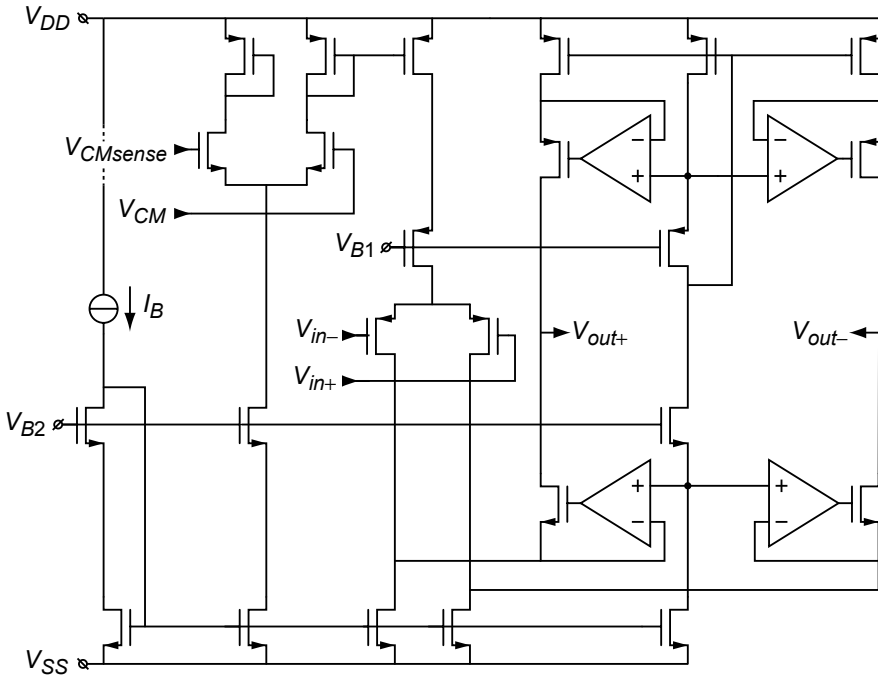


Figure 7.26. Circuit diagram of the gain-boosted folded-cascode opamp used in the first integrator.

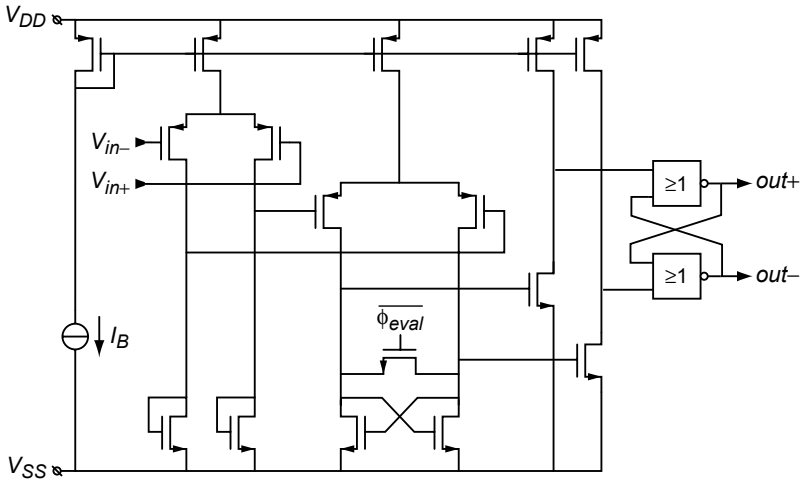


Figure 7.27. Circuit diagram of the latched comparator.

was therefore designed for complete settling (i.e. settling errors equivalent to less than 0.01 K). The tail current of the opamp was chosen larger ($3\ \mu\text{A}$) than the smallest bias current used for generating ΔV_{BE} ($1\ \mu\text{A}$), so that the settling behaviour is determined by that bias current, rather than by the opamp. To guarantee negligible leakage in the first integrator, and to ensure that errors introduced after the first integrator are negligible, gain-boosting has been used to get a DC gain well above 100 dB under all operating conditions and process corners [10, 11]. A capacitive divider (not shown) is used to sense the common-mode voltage at the output of the first integrator.

Since errors introduced by the second integrator are attenuated by the gain of the first integrator, no offset cancellation, DEM, or gain boosting are needed here. A simple folded-cascode opamp is used. The comparator is implemented as a dynamic latch preceded by a preamp, which prevents kickback to the output of the second integrator (Figure 7.27).

The modulator uses non-overlapping clocks. Switching in the front-end circuitry and updating of the bitstream output take place in the time gap between the clock phases ϕ_1 and ϕ_2 , so that any resulting charge injection does not result in errors. Clocks with delayed falling edges (e.g. ϕ_{F1d}) are used to prevent signal-dependent charge injection [4].

Offset Cancellation

While offset and $1/f$ noise of the first integrator are reduced by the applied autozeroing, charge-injection mismatch in the switches in the first integrator results in residual offset (see Section 5.3.5). Minimum-size NMOS switches are used to minimize this offset. Nevertheless, an offset of a few tens of μV remains. This is too large, given that the error in ΔV_{BE} has to be in the order of $2\ \mu\text{V}$.

To further reduce the offset, the modulator is chopped at system-level. A chopper switch at the input and a switch at the output periodically reverse the polarity of the input signal and the bitstream. To avoid disturbing the operation of the modulator when chopping, its state is also inverted by swapping the integration capacitors of both integrators. The chopping is done at a slow speed to ensure that errors due to charge injection in the chopper switches are negligible. Two chopping periods per conversion are used, so as to modulate the offset to the first zero of the decimation filter (see Figure 7.29).

Clock Boosting

To make sure the minimum-size NMOS switches in the first integrator are properly turned on even at the lowest supply voltage of 2.5 V, they are driven by boosted clock signals V_{sw1} and V_{sw2} , the timing of which corresponds to the regular clock signals ϕ_1 and ϕ_2 , respectively. The clock-boosting circuit used for generating the drive signals V_{sw1} and V_{sw2} is shown in Figure 7.28. It is

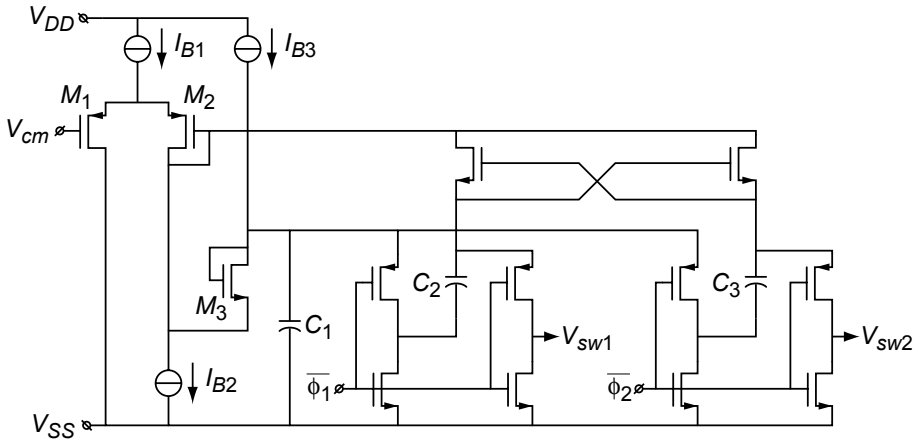


Figure 7.28. Clock-boosting circuit for driving the NMOS switches in the first integrator.

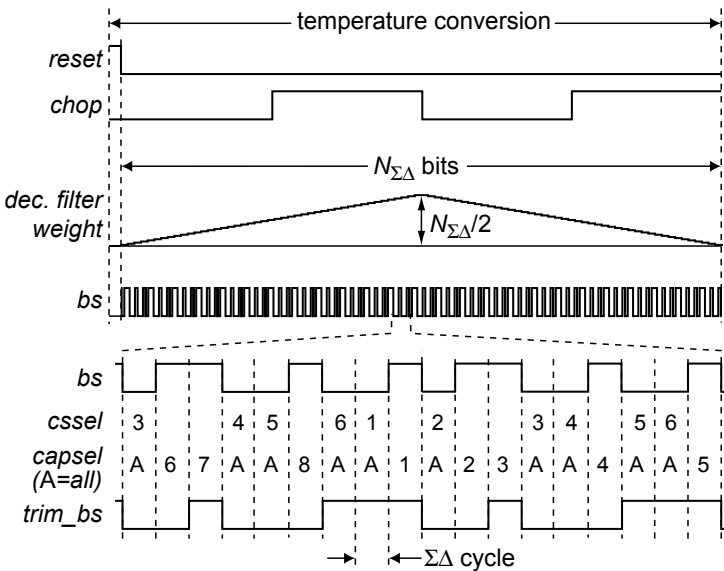


Figure 7.29. Timing diagram of a temperature conversion, showing the system-level chopping, the impulse response of the decimation filter, and a fragment of the bitstream with DEM control signals for the current sources (*cssel*) and the sampling capacitors (*capsel*), and the trimming bitstream for fine-trim setting $F = 50\%$.

similar to the clock-doubling circuits often used in switched-capacitor designs [12], except that the drive signals generated by this circuit are, to first order, independent of the supply voltage. This makes the charge injection of the switches also less supply dependent.

A voltage follower consisting of M_1 and M_2 produces a buffered version of the modulator's common-mode reference $V_{cm} \simeq 1.3\text{ V}$. When $\phi_1 = 0$, capacitor C_2 is charged to V_{cm} , while the output V_{sw1} is pulled to ground, thus switching off $M_{1a,b}$ in Figure 7.25. When $\phi_1 = 1$, the bottom plate of C_2 is lifted to $V_{cm} + V_{gs3}$. As a result, the output V_{sw1} rises to $2V_{cm} + V_{gs3} \simeq 3.3\text{ V}$. As the gates $M_{1a,b}$ in Figure 7.25 are connected to the virtual ground of the first opamp, which has a common-mode level equal to V_{cm} , their gate-source voltage becomes $V_{cm} + V_{gs3}$, turning them fully on. Drive signal V_{sw2} is generated similarly using control signal ϕ_2 and capacitor C_3 . Capacitor C_1 serves as a buffer which provides charge when C_2 or C_3 is lifted up.

7.3.7 Timing and Decimation Filter

An overview of the timing of a complete temperature conversion is shown in Figure 7.29. A conversion starts with a reset of both integrators of the $\Sigma\Delta$ modulator in order to bring the modulator into a well-defined state. After that, it runs for a number of clock cycles, producing a bitstream of $N_{\Sigma\Delta}$ bits, which are processed by a sinc^2 decimation filter (with a symmetrical triangular impulse response). The system-level chopping is done twice per conversion, so that the offset is modulated to the frequency of the first zero of this filter.

An arbitrary fragment of the bitstream is shown in Figure 7.29 to illustrate the timing of the DEM of the 1 : 5 current ratio and the 1 : 8 sampling capacitor ratio, and the timing of the $\Sigma\Delta$ -modulated trimming of the bias current. The bitstream-controlled timing introduced in Section 4.6.2 is used to prevent errors due to intermodulation between the bitstream and residuals of the DEM and the modulated trimming.

The current source used for generating the unit current in the 1 : 5 ratio is selected by a cyclic counter that counts from 1 to 6. This counter is enabled only if the bitstream is 0. Similarly, the unit capacitor used for sampling V_{BE} is selected by a cyclic 1 to 8 counter, which is enabled only if the bitstream is 1 [6].

The digital $\Sigma\Delta$ modulator that produces the fine-trimming bitstream $trim_bs$ (Figure 7.22), is only clocked at the end of $\Sigma\Delta$ cycles in which $bs = 1$. If $bs = 0$, the modulator is frozen. Figure 7.29 shows how this works out if $trim_bs$ has a 50% duty cycle: a repetitive 0101 pattern appears in successive $\Sigma\Delta$ cycles in which $bs = 1$.

This 'bitstream-controlled' operation of the digital $\Sigma\Delta$ modulator prevents quantization noise from being modulated into the signal band due to intermodulation between the two bitstreams [8]. Such intermodulation can occur, because

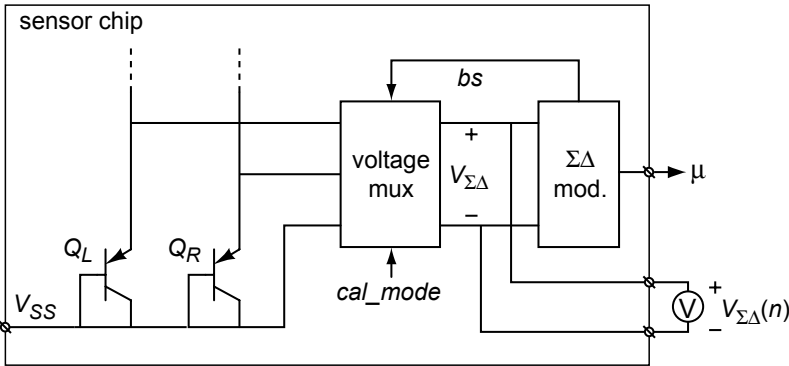


Figure 7.30. Configuration for calibrating the sensor by measuring the on-chip ΔV_{BE} .

the two bitstreams are effectively multiplied; this is a result of the fact that V_{BE} , which is modulated by $trim_bs$, is only integrated when $bs = 1$.

7.3.8 Calibration

As mentioned before, the sensor needs to be trimmed to compensate for spread of the base-emitter voltage V_{BE} of the substrate bipolar transistors. A calibration procedure is used to determine, directly or indirectly, how much this voltage differs from the desired value, so that appropriate trim settings (C and F in Figure 7.22) can be found. This is done after packaging, in order to take stress-related shifts into account. All calibration techniques described in Chapter 6 can be applied to this sensor: conventional calibration using a reference thermometer, batch calibration, calibration based on ΔV_{BE} measurement, and voltage reference calibration. The implementation of the latter two techniques requires some further explanation.

Calibration based on ΔV_{BE} Measurement

The configuration used for calibration based on ΔV_{BE} measurement is shown in Figure 7.30. The input voltage of the $\Sigma\Delta$ modulator is output via two pins and measured using an accurate external voltmeter. The voltage multiplexer is configured to successively output ΔV_{BE} for all combinations of:

- $cssel \in \{1, 2, \dots, 6\}$: the 6 DEM steps of the 1 : 5 current sources,
- $\phi_L \in \{0, 1\}$: the larger current going through Q_L or through Q_R ,
- $\phi_{high} \in \{0, 1\}$: nominal currents of $1 \mu\text{A}$ or $3 \mu\text{A}$ (this is a test mode built into the bias circuit).

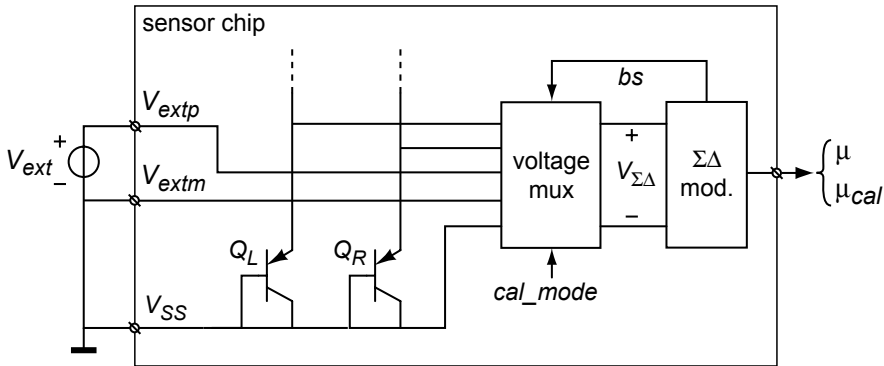


Figure 7.31. Configuration for calibrating the sensor by applying an external voltage V_{ext} .

From the resulting 24 voltages the intrinsic ΔV_{BE} can be calculated, while compensating for offsets (both between the transistors and in the readout chain), for mismatch in the current ratio, and for series resistance (see Section 3.7.3). From this, in turn, the chip's temperature can be calculated, which can then be used to trim the chip.

This approach relies on the intrinsic accuracy of ΔV_{BE} . Effectively, any error sources *except* those in ΔV_{BE} can be corrected for. A disadvantage is, however, that small on-chip voltages have to be measured accurately (approximately 40 mV with an inaccuracy below $10 \mu\text{V}$). In that respect, the technique described next may be a more attractive alternative in production environment.

Voltage Reference Calibration

Voltage reference calibration is based on applying an accurate external reference voltage to the chip, rather than on reading out small on-chip voltages. This is illustrated in Figure 7.31. For simplicity of implementation, the external voltage V_{ext} replaces the on-chip V_{BE} during calibration, rather than acting as a true reference for the ADC (which would require a different charge balancing scheme than in normal operation). When V_{BE} is replaced by an external voltage V_{ext} , the bitstream average becomes

$$\mu_{cal} = \frac{16 \cdot \Delta V_{BE}}{V_{ext} + 16 \cdot \Delta V_{BE}}, \quad (7.13)$$

which makes it possible to determine ΔV_{BE} , and hence the chip's temperature, from the applied voltage V_{ext} and the measured average μ_{cal} . This indirectly measured ΔV_{BE} can then, again, be used to trim the chip.

An essential difference with calibration based on direct ΔV_{BE} measurement is, that this approach relies on the accuracy of *all* on-chip circuitry. Errors except those resulting from V_{BE} spread will therefore result in residual temperature

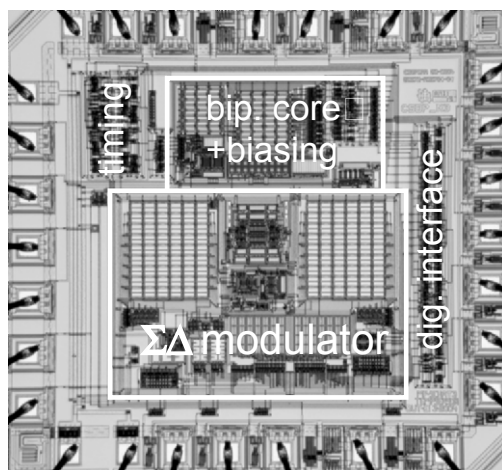


Figure 7.32. Micrograph of the temperature sensor testchip.

errors after trimming, not only towards the ends of the temperature range, but also at the calibration temperature. Provided that all such errors have been reduced to negligible levels by design, this is not a problem. The main advantage is that the voltage to be applied to the chip is much larger than the voltages that have to be measured in the case of calibration based on direct ΔV_{BE} measurement (approximately 600 mV with an inaccuracy below 0.2 mV). This makes voltage reference calibration much more robust against interference.

7.3.9 Experimental Results

The sensor has been realized in a 0.7 μm CMOS process of AMI Semiconductor with linear capacitors and high-resistivity poly resistors. A chip micrograph is shown in Figure 7.32. The chip area is 4.5 mm², which includes bondpads and some test circuitry. The decimation filter and digital control circuitry were implemented off-chip for testing flexibility. A version of the chip with on-chip serial interface and control circuitry was also realized (Figure 7.33).

Figure 7.34 shows measured power spectra of the bitstream of the second-order $\Sigma\Delta$ modulator. The second-order noise shaping is clearly visible. The figure shows the effectiveness of the bitstream-controlled operation of the digital trimming modulator. With a free-running trimming modulator, the noise floor increases by about 30 dB as a result of intermodulated quantization noise that ends up in the signal band. A similar increase in the noise floor is observed when the counters that control the DEM of the sampling capacitors and the current sources are free-running rather than bitstream-controlled [6]. The noise

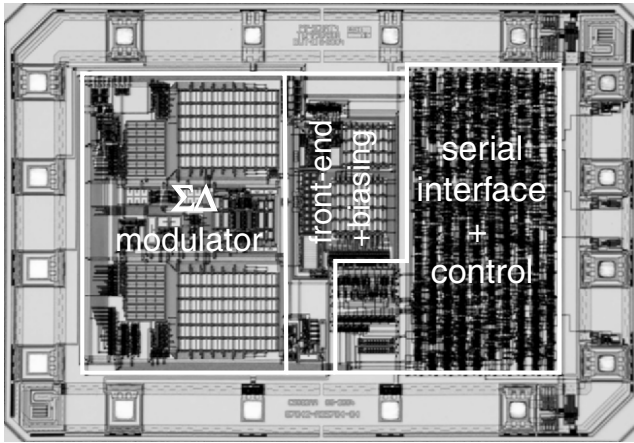


Figure 7.33. Micrograph of the temperature sensor testchip with on-chip digital circuitry.

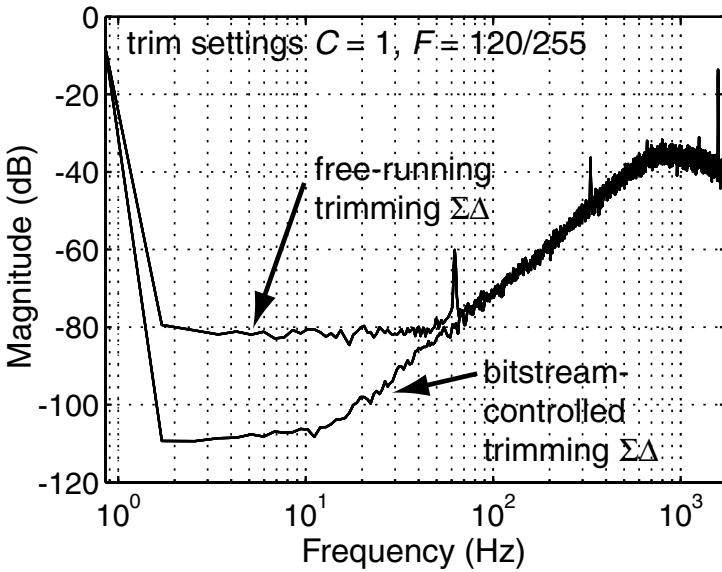


Figure 7.34. Measured power spectrum of the bitstream, with the digital trimming $\Sigma\Delta$ modulator operated in free-running and in bitstream-controlled mode (4096 bits, 16x averaged, Hanning windowed).

floor measured with bitstream-controlled timing is identical to that measured with the DEM and the trimming modulator disabled. This shows that in-band intermodulation products are indeed completely eliminated.

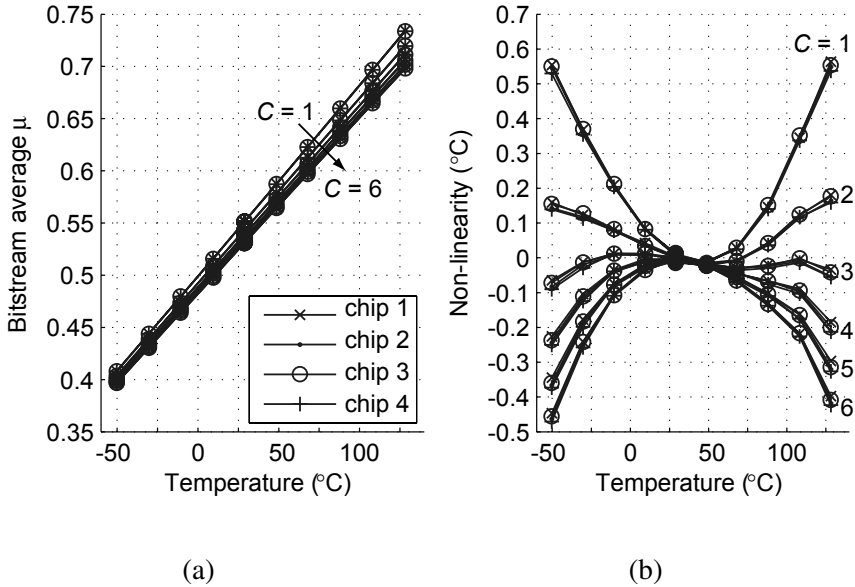


Figure 7.35. (a) Measured bitstream averages of four devices as a function of temperature for the coarse trim settings $C = 1$ to $C = 6$; (b) non-linearity of these bitstream averages.

To characterize the sensor, 24 samples from one batch were mounted in ceramic packages, placed in an oven, and compared with a platinum thermometer. This thermometer was calibrated to 20 mK at the Dutch Metrology Institute.

First, the bitstream averages of four samples were measured in the temperature range from -55°C to 125°C in order to characterize the sensor's non-linearity. The coarse-trim setting of the bias current (i.e. input C in Figure 7.22) was varied between 1 and 6. The measured bitstream averages are shown in Figure 7.35a. Figure 7.35b shows their non-linearity expressed in $^{\circ}\text{C}$. For a given trim setting, the differences between the four devices are small. The lowest trim setting $C = 1$ causes, as expected, the highest bitstream averages, and results in a convex second-order non-linearity. If the trim setting is increased, a larger bias current is used for generating V_{BE} . This effectively increases the $\Sigma\Delta$ modulator's reference, and therefore results in a decrease of the bitstream average. It also increases the temperature coefficient of the reference, which, in agreement with the ratiometric curvature correction technique discussed in Section 3.5.4, results in a decrease in non-linearity. For $C = 3$, a minimum residual non-linearity of about 0.1°C is obtained. Its shape is in close agreement with that shown in Figure 3.24, except that it slightly bends down at the high end of the temperature range. This is probably due to leakage currents

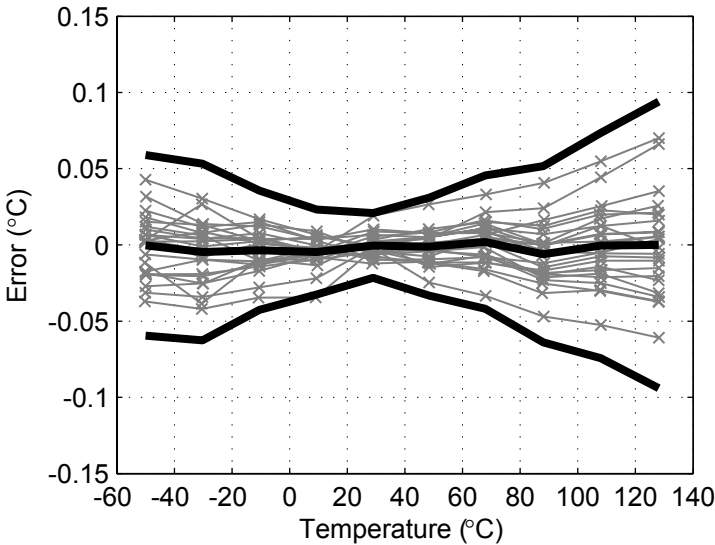


Figure 7.36. Measured temperature error of 24 devices after calibration at 30°C using a Pt100 thermometer; bold lines indicate the average error and $\pm 3\sigma$ values.

in the first integrator of the $\Sigma\Delta$ modulator. The residual non-linearity can be eliminated using a non-linear decimation filter (see Section 4.5.4).

The measured bitstream averages of one of the four devices was used to establish the values of scaling parameters A and B needed to convert a bitstream average μ into a temperature reading D_{out} in $^\circ\text{C}$ using equation (3.7): $D_{out} = A \cdot \mu + B$. For the trim setting that results in minimum non-linearity, $C = 3$, these parameters were $A = 574.24^\circ\text{C}$ and $B = 280.6^\circ\text{C}$. Assuming that PTAT spread of V_{BE} is the only significant source of errors, the other samples should have the same minimum non-linearity after trimming if the same values for A and B are used. Therefore, the bitstream outputs of all sensors were processed with a decimation filter (in software) that uses these same scaling parameters. This decimation filter also compensates for the residual non-linearity.

Figure 7.36 shows the measured temperature error of all 24 same samples after trimming based on a conventional calibration. The coarse and fine trimming parameters C and F (see Figure 7.22) of every individual sensor were adjusted so as to null the temperature error at 30°C , which was determined by comparison with the platinum thermometer. After trimming, the samples have a 3σ inaccuracy of only $\pm 0.1^\circ\text{C}$ over the full military temperature range. This performance confirms the above-mentioned assumption that PTAT spread

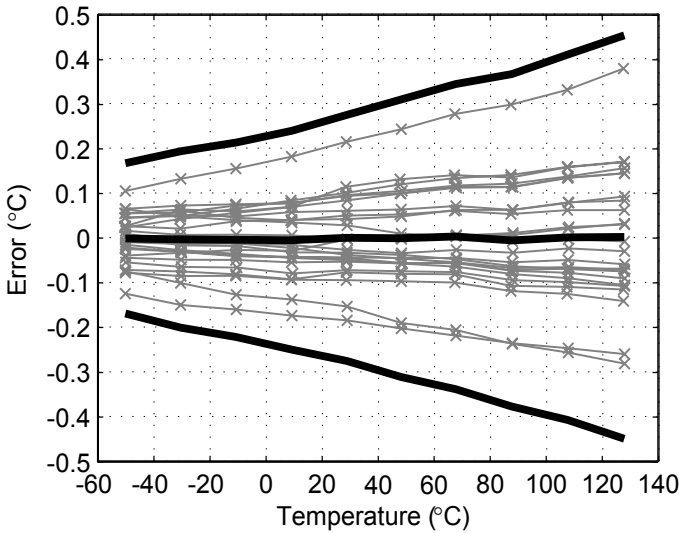


Figure 7.37. Measured temperature error of 24 devices after batch calibration; bold lines indicate the average error and $\pm 3\sigma$ values.

of V_{BE} is the only significant error source, and shows the effectiveness of the applied readout techniques. It is the highest reported accuracy to date.

The high accuracy obtained using the conventional calibration technique comes at the cost of a long calibration time. A cheaper alternative is to use batch calibration. This boils down to using the same trim setting for all devices from one batch. This trim settings is determined by calibrating a small number of samples from the batch using the conventional calibration technique. Figure 7.37 shows the temperature error if the average trim setting of the first four devices is used for all 24 devices. A 3σ inaccuracy of $\pm 0.5^\circ\text{C}$ over the military temperature range is obtained, which corresponds to the intra-batch spread of this sensor. As a result of the extra precision techniques applied, the performance is three times better than that of the batch-calibrated sensor discussed in Section 7.1.

Another low-cost alternative to conventional calibration is voltage reference calibration. Figure 7.38 shows the measured temperature error of 16 devices that have been calibration using that technique. As described in Section 7.3.8, an external voltage of 600 mV was applied to the devices in order to indirectly measure the on-chip ΔV_{BE} . This was then used to calculate the temperature and trim the devices. The 3σ inaccuracy obtained over the military temperature range is $\pm 0.25^\circ\text{C}$. While significantly better than that obtained using batch calibration, it is less than expected, because all interface electronics were designed for an inaccuracy of $\pm 0.1^\circ\text{C}$. An explanation for this unexpected

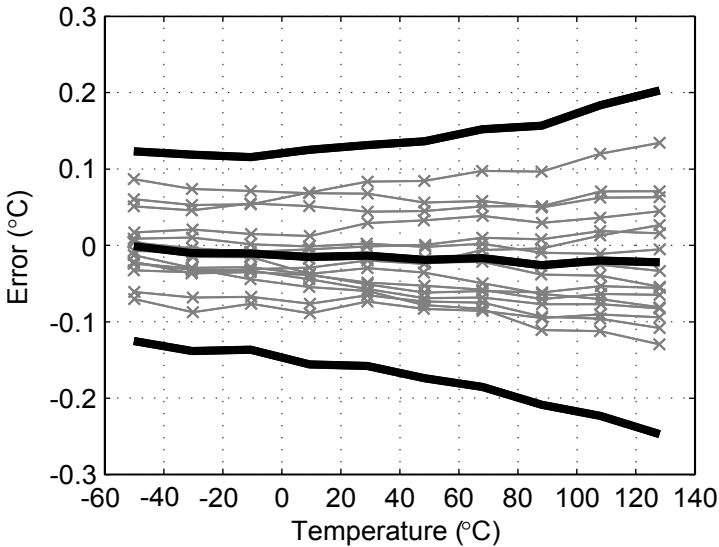


Figure 7.38. Measured temperature error of 16 devices after voltage reference calibration; bold lines indicate the average error and $\pm 3\sigma$ values.

inaccuracy was found in the layout of the chip: a small parasitic interconnect capacitor of about 15 fF is presented in parallel with the sampling capacitors of the $\Sigma\Delta$ modulator. This capacitor causes a gain error that is not eliminated by means of dynamic element matching. In the case of conventional calibration, this gain error is trimmed out at the calibration temperature, and causes some residual errors towards the end of the temperature range, which is consistent with the results shown in Figure 7.36. Voltage reference calibration, in contrast, relies on the accuracy of all readout circuitry, and therefore performs worse. The parasitic capacitor will be removed in a redesign.

Calibration based on ΔV_{BE} measurement does not rely on the intrinsic accuracy of the $\Sigma\Delta$ modulator, and should therefore perform better. As described in Section 7.3.8, a sequence of base-emitter voltages is measured and used to calculate ΔV_{BE} . From that, the device temperature is calculated. Figure 7.39 shows the measured temperature error of all 24 devices after calibration based on that temperature. The 3σ inaccuracy over the military temperature range is $\pm 0.15^\circ\text{C}$. A small systematic error is visible which can be attributed to inaccuracy in the characterization of the sensitivity of ΔV_{BE} . These results show that a high accuracy can be obtained using a low-cost calibration technique.

All above-mentioned measurements were performed at a supply voltage of 3.3 V. The sensor is functional for supply voltages from 2.5 V to 5.5 V. Over this range, the power-supply sensitivity is $0.03^\circ\text{C}/\text{V}$, which is a significant

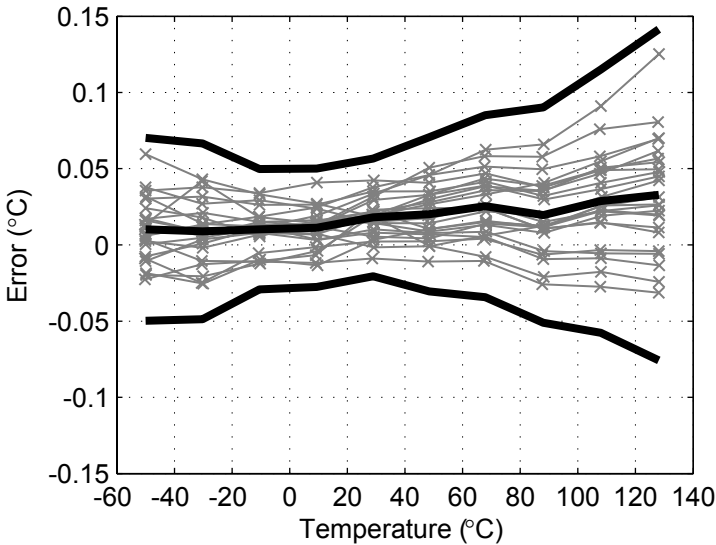


Figure 7.39. Measured temperature error of 24 devices after calibration based on ΔV_{BE} measurement; bold lines indicate the average error and $\pm 3\sigma$ values.

Table 7.5. Performance summary of the temperature sensor.

Parameter	Value	Conditions
Process	0.7 μm CMOS	
Area	4.5 mm^2	
Inaccuracy ($\pm 3\sigma$)	$\pm 0.5^\circ\text{C}$ $\pm 0.25^\circ\text{C}$ $\pm 0.15^\circ\text{C}$ $\pm 0.1^\circ\text{C}$ $\pm 0.03^\circ\text{C}$	-55°C to 125°C , batch cal. -55°C to 125°C , voltage ref. cal. -55°C to 125°C , ΔV_{BE} meas. cal. -55°C to 125°C , Pt100 cal. at 30°C , Pt100 cal.
Resolution	0.01°C	100 ms conversion time
Supply voltage	2.5 V – 5.5 V	
Supply sensitivity	$0.03^\circ\text{C}/\text{V}$	
Supply current	75 μA	continuous operation

improvement over the previous designs. This can be attributed to the supply-insensitive bias circuit and the fully differential circuitry. A performance summary is given in Table 7.5.

Table 7.6. Accuracy comparison of smart temperature sensors.

Reference	Inaccuracy	Range	Conditions	Calibration / Trimming
Bakker, 1996 [2, 13]	$\pm 1.0^\circ\text{C}$	-40°C to 120°C	min/max of 3 samples	2 points, after packaging
Tuthill, 1998 [14]	$\pm 1.5^\circ\text{C}$	-50°C to 125°C	min/max of 6 samples	1 point, wafer-level
Bakker, 1999 [2, 15]	$\pm 2.0^\circ\text{C}$	-20°C to 100°C	min/max 112 samples	no trimming
Hagleitner, 2002 [16]	$\pm 0.3^\circ\text{C}$	-40°C to 120°C	min/max	full-range linear fit
LM75 [1]	$\pm 2.0^\circ\text{C}$	-25°C to 100°C	min/max	unknown
LM92 [17]	$\pm 3.0^\circ\text{C}$	-55°C to 125°C	min/max	unknown
DS1626 [18],	$\pm 0.33^\circ\text{C}$	at 30°C	min/max	unknown
ADT7301 [19]	$\pm 1.5^\circ\text{C}$	-25°C to 150°C	min/max	unknown
SMT160-30 [20]	$\pm 0.5^\circ\text{C}$	0°C to 70°C	min/max	unknown
	$\pm 2.0^\circ\text{C}$	-55°C to 125°C	min/max	unknown
	$\pm 1.0^\circ\text{C}$	0°C to 70°C	min/max	unknown
	$\pm 3.0^\circ\text{C}$	-40°C to 125°C	min/max	unknown
	$\pm 0.7^\circ\text{C}$	-30°C to 100°C	min/max	unknown
	$\pm 1.2^\circ\text{C}$	-45°C to 130°C	min/max	unknown
This work, Section 7.1 [21]	$\pm 1.5^\circ\text{C}$	-55°C to 120°C	$\pm 3\sigma$ of 50 samples	batch calibration
This work, Section 7.2 [22]	$\pm 0.3^\circ\text{C}$	at 25°C	$\pm 3\sigma$ of 32 samples	cal. based on ΔV_{BE} meas.
	$\pm 0.5^\circ\text{C}$	-50°C to 120°C	$\pm 3\sigma$ of 24 samples	batch calibration
This work, Section 7.3 [23]	$\pm 0.5^\circ\text{C}$	-55°C to 125°C	$\pm 3\sigma$ of 16 samples	voltage reference cal.
	$\pm 0.25^\circ\text{C}$	-55°C to 125°C	$\pm 3\sigma$ of 24 samples	cal. based on ΔV_{BE} meas.
	$\pm 0.15^\circ\text{C}$	-55°C to 125°C	$\pm 3\sigma$ of 24 samples	} cal. using Pt100
	$\pm 0.03^\circ\text{C}$	at 30°C	$\pm 3\sigma$ of 24 samples	
	$\pm 0.1^\circ\text{C}$	-55°C to 125°C	$\pm 3\sigma$ of 24 samples	

7.4 Benchmark

Table 7.6 compares the performance of the sensors discussed in this chapter with previous work. Since most work in the field of smart temperature sensors is done in industry, the specifications of five leading commercial sensors have also been included. This table shows that all sensors described in this chapter have state-of-the-art accuracy. The sensor described in Section 7.3 achieves the highest reported accuracy to date.

References

- [1] “LM75 data sheet,” National Semiconductor Corp., Feb. 2004, www.national.com.
- [2] A. Bakker and J. H. Huijsing, *High-Accuracy CMOS Smart Temperature Sensors*. Boston: Kluwer Academic Publishers, 2000.
- [3] H. Yoshizawa, Y. Huang, P. F. Ferguson, and G. C. Temes, “MOSFET-only switched-capacitor circuits in digital CMOS technology,” *IEEE Journal of Solid-State Circuits*, vol. 34, no. 6, pp. 734–747, June 1999.
- [4] S. R. Norsworthy, R. Schreier, and G. C. Temes, Eds., *Delta-Sigma Data Converters: Theory, Design and Simulation*. Piscataway, New York: IEEE Press, 1997.
- [5] C. Hagleitner *et al.*, “A gas detection system on a single CMOS chip comprising capacitive, calorimetric, and mass-sensitive microsensors,” in *Dig. Techn. Papers ISSCC*, Feb. 2002, pp. 430–431, 479.
- [6] M. A. P. Pertijs and J. H. Huijsing, “A sigma-delta modulator with bitstream-controlled dynamic element matching,” in *Proc. ESSCIRC*, Sept. 2004, pp. 187–190.
- [7] G. C. M. Meijer, G. Wang, and F. Fruett, “Temperature sensors and voltage references implemented in CMOS technology,” *IEEE Sensors Journal*, vol. 1, no. 3, pp. 225–234, Oct. 2001.
- [8] M. A. P. Pertijs and J. H. Huijsing, “Bitstream trimming of a smart temperature sensor,” in *Proc. IEEE Sensors*, Oct. 2004, pp. 904–907.
- [9] G. v. d. Horn and J. H. Huijsing, *Integrated Smart Sensors: Design and Calibration*. Boston: Kluwer Academic Publishers, 1998.
- [10] K. Bult and G. J. G. M. Geelen, “A fast-settling CMOS op amp for SC circuits with 90-dB DC gain,” *IEEE Journal of Solid-State Circuits*, vol. 25, pp. 1379–1384, Dec. 1990.
- [11] J. H. Huijsing, R. Hogervorst, and K. de Langen, “Low-power low-voltage VLSI operational amplifier cells,” *IEEE Transactions on Circuits and Systems—Part I: Fundamental Theory and Applications*, vol. 42, pp. 841–852, Nov. 1995.
- [12] S. Rabii and B. A. Wooley, *The Design of Low-Voltage, Low-Power Sigma-Delta Modulators*. Boston: Kluwer Academic Publishers, 1999.
- [13] A. Bakker and J. H. Huijsing, “Micropower CMOS temperature sensor with digital output,” *IEEE Journal of Solid-State Circuits*, vol. 31, no. 7, pp. 933–937, July 1996.

- [14] M. Tuthill, "A switched-current, switched-capacitor temperature sensor in 0.6- μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 7, pp. 1117–1122, 1998.
- [15] A. Bakker and J. H. Huijsing, "A low-cost high-accuracy CMOS smart temperature sensor," in *Proc. ESSCIRC*, Sept. 1999, pp. 302–305.
- [16] C. Hagleitner, "CMOS single-chip gas detection system comprising capacitive, calorimetric and mass-sensitive microsensors," Ph.D. dissertation, Swiss Federal Institute of Technology, Zurich, Switzerland, 2002.
- [17] "LM92 data sheet," National Semiconductor Corp., Mar. 2005, www.national.com.
- [18] "DS1626 data sheet," Maxim Int. Prod., May 2005, www.maxim-ic.com.
- [19] "ADT7301 data sheet," Analog Devices Inc., Aug. 2004, www.analog.com.
- [20] "SMT160-30 data sheet," Smartec B.V., May 2003, www.smartec.nl.
- [21] M. A. P. Pertijs, A. Bakker, and J. H. Huijsing, "A high-accuracy temperature sensor with second-order curvature correction and digital bus interface," in *Proc. ISCAS*, May 2001, pp. 368–371.
- [22] M. A. P. Pertijs, A. Niederkorn, X. Ma, B. McKillop, A. Bakker, and J. H. Huijsing, "A CMOS smart temperature sensor with a 3σ inaccuracy of $\pm 0.5^\circ\text{C}$ from -50°C to 120°C ," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 2, pp. 454–461, Feb. 2005.
- [23] M. A. P. Pertijs, K. A. A. Makinwa, and J. H. Huijsing, "A CMOS smart temperature sensor with a 3σ inaccuracy of $\pm 0.1^\circ\text{C}$ from -55°C to 125°C ," *IEEE Journal of Solid-State Circuits*, in press.

Chapter 8

CONCLUSIONS

This final chapter summarizes the main findings of this book. It also shows that some of the techniques developed for smart temperature sensors can also be applied to other applications, and provides an outlook on future work on CMOS smart temperature sensors.

8.1 Main Findings

The following list summarizes the main findings of this book:

- Substrate pnp transistors are the device of choice for the implementation of precision CMOS smart temperature sensors (Chapter 2)¹.
- Using precision biasing and readout techniques, low-stress packaging and a single-temperature calibration, temperature errors due to the non-idealities of substrate pnp transistors can be reduced to less than ± 0.1 °C (Chapter 3).
- Residual errors are then determined by packaging stress, and by spread of the curvature of the base-emitter voltage. The latter can result from spread of the temperature dependency of the saturation current, and/or from spread of the temperature dependency of the bias resistor (Chapter 3).
- Sigma-delta ADCs are the most suitable ADCs for smart temperature sensors. These ADCs can provide a readily interpretable digital output, and can take care of the filtering of dynamic error signals produced by the precision techniques applied in the front-end circuitry (Chapter 4).
- For precision CMOS smart temperature sensors, a switched-capacitor implementation is preferred over a continuous-time implementation in view of

¹This conclusion has also been drawn in previous work [1, 2].

its compatibility with dynamic error correction techniques. In most standard packages, the somewhat higher power consumption of a switched-capacitor implementation need not result in significant self-heating (Chapter 5).

- Calibration after packaging can be implemented in a cost-effective way if it is based on voltage measurements rather than temperature measurements (Chapter 6).
- It is possible to achieve a 3σ inaccuracy of only ± 0.1 °C over the full military temperature range with a CMOS smart temperature sensor that is packaged in ceramic and calibrated only at room temperature (Chapter 7).

8.2 Other Applications of this Work

While this work has focused on temperature sensors that measure their own temperature, many commercial sensors offer the possibility to also measure the temperature of one or more external diodes (see, for instance, [3]). This feature is used in thermal management systems in PCs and laptops, where the temperature of a microprocessor is measured using a diode on the same substrate. The measured temperature is used, for instance, to regulate the operation of a cooling fan. The inaccuracy of such systems, which is typically ± 3 °C, can possibly be reduced using the techniques described in this work. For instance, the techniques described in Section 3.7.3 can be applied to eliminate errors due to series resistances associated with the remote diode.

The techniques described in this work can also be applied in voltage measurement systems. In such systems, an external voltage has to be compared to an accurate reference. This is similar to the comparison of ΔV_{BE} to an accurate dynamic reference in a smart temperature sensor. The various described dynamic error correction techniques can therefore also be applied in this application. Thus, potentially, a higher accuracy can be obtained than with a stand-alone voltage reference and a separate ADC. Moreover, if a single chip can digitize both temperature and an external voltage, it is easy to implement compensation for temperature cross-sensitivity.

The circuit techniques needed in a precision CMOS temperature sensor may also be interesting for other applications, such as interfaces for other types of sensors and data acquisition circuits. The bitstream-controlled timing described in Section 4.6.2, for instance, can be applied whenever dynamic error correction techniques, such as dynamic element matching, are applied in the front-end of a sigma-delta ADC.

8.3 Future Work

The following topics would be interesting to address in future work on CMOS smart temperature sensors:

- Incorporation of a stress sensor to compensate for errors due to packaging stress. As explained in Sections 2.6 and 3.4.3, the sensitivity of bipolar transistors to (temperature-dependent) packaging stress limits the accuracy that can be obtained. This is especially true in low-cost plastic packages, even if trimming is performed after packaging. Perhaps some compensation can be obtained using an on-chip (ratiometric) stress sensor. The sensors proposed in [4] could be used for this purpose. A stress measurement at wafer-level could be performed, the result of which could be stored in non-volatile memory. During operation, the sensor could then compare the current stress to the stored value, and correct the measured temperature accordingly.
- Further investigation of the reverse Early effect. As explained in Section 2.7.3, the reverse Early effect leads to a multiplicative error in the base-emitter voltage. This error is often modelled using a non-unity effective emission coefficient n_F . As mentioned in Section 2.7.3, this error cancels in a ratiometric temperature sensor, because it affects both V_{BE} and ΔV_{BE} in the same way. However, significant measurement errors may result if temperature is determined based on a measurement of only ΔV_{BE} , using a voltage reference that is not affected by the same multiplicative error. This is the case for the calibration techniques proposed in Chapter 6, and also for the remote-diode measurement mentioned in Section 8.2. These applications rely on the reproducibility of the reverse Early effect. Since little is known about this reproducibility (at least in the open literature), further investigation would be useful. Also, it would be interesting to investigate if it is possible to extract the reverse Early voltage from (simple) measurements, so that compensation for the multiplicative error becomes possible. A measurable parameter that could make such an extraction possible is the forward current-gain, which is affected by the reverse Early effect in a similar way as the collector current [5].
- Further investigation into the effect of spread of the bias resistor. As explained in Section 2.8.3, spread of the temperature coefficient (TC) of the bias resistor leads to a non-PTAT spread of the base-emitter voltage. Such spread cannot be trimmed out based on a single-temperature calibration, and is therefore one of the factors that limit the accuracy. It would be interesting to investigate if it is possible to exploit the correlation between the TC of the bias resistor and its absolute value at the calibration temperature. An estimate of the TC obtained from a resistance measurement could be used to trim out the non-PTAT spread.
- Further exploration of the trade-off between power consumption and accuracy. Low-power temperature sensors are needed in battery-operated and

wireless applications [6]. In Chapter 5, power consumption has been analyzed mainly to verify that errors related to self-heating are small enough. A more detailed study is needed to find the absolute minimum power consumption needed to achieve a given accuracy. A continuous-time topology is then probably preferred. For ultra-low-power applications a non-oversampling ADC is probably needed (see e.g. [7]).

- Development of a compact temperature sensing module. Such a module would be interesting for use in smart sensors to compensate for temperature cross-sensitivity, and for thermal management in power devices. While such applications typically have modest accuracy requirements, it would be interesting to explore how the described dynamic error correction techniques can help to achieve this modest accuracy with a much smaller chip area. In a switched-capacitor design, for instance, a modest accuracy might be achievable without using dynamic element matching of the sampling capacitors. However, the same accuracy can probably be achieved with much smaller capacitors if dynamic element matching is applied.
- Development of a temperature sensor in deep sub-micron CMOS. Such a sensor is useful for thermal management of modern microprocessors and their peripheral ICs. With the very poor performance of the bipolar transistors available in these processes, an altogether different sensing principle may be required, even to obtain the modest accuracy needed in this application.

References

- [1] A. Bakker and J. H. Huijsing, *High-Accuracy CMOS Smart Temperature Sensors*. Boston: Kluwer Academic Publishers, 2000.
- [2] G. Wang and G. C. M. Meijer, "Temperature characteristics of bipolar transistors fabricated in CMOS technology," *Sensors and Actuators*, vol. 87, pp. 81–89, Dec. 2000.
- [3] "NE1617A data sheet," Philips Semiconductors, Oct. 2004, www.semiconductors.philips.com.
- [4] F. Fruett and G. C. M. Meijer, *The Piezjunction Effect in Silicon Integrated Circuits and Sensors*. Boston: Kluwer Academic Publishers, May 2002.
- [5] I. E. Getreu, *Modeling the Bipolar Transistor*. Amsterdam, The Netherlands: Elsevier, 1976.
- [6] A. Bakker and J. H. Huijsing, "Micropower CMOS temperature sensor with digital output," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 7, pp. 933–937, July 1996.
- [7] M. D. Scott, B. E. Boser, and K. S. J. Pister, "An ultralow-energy ADC for smart dust," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 7, pp. 1123–1129, July 2003.

Appendix A

Derivation of Mismatch-Related Errors

A.1 Errors in ΔV_{BE}

To generate ΔV_{BE} , bias currents with a well-defined ratio $1 : p$ are required. These are usually generated using a set of $p + 1$ current sources. Mismatch between these current sources results in an error in ΔV_{BE} . In the first part of this section, an expression for this error will be derived as a function of the mismatch between the current sources.

Dynamic element matching (DEM) can be used to reduce the error that results from current-source mismatches. Since DEM does not completely eliminate the error, it is important to estimate the maximum error that remains if DEM is applied. An estimate of that maximum error, as derived in the second part of this section, can be used to determine the initial matching required to obtain a given performance.

A.1.1 Without DEM

Lets assume that ΔV_{BE} is generated using a $1 : p$ bias current ratio realized using $p + 1$ current sources with values

$$I_i = I(1 + \delta_i), \quad 1 \leq i \leq p + 1, \quad (\text{A.1})$$

where δ_i is the relative error of the i^{th} current source as a result of mismatch with respect to average current I , and hence

$$\sum_{i=1}^{p+1} \delta_i = 0. \quad (\text{A.2})$$

If the j^{th} current source is used to generate the unit current, the resulting current ratio is

$$\frac{\sum_{i \neq j} I_i}{I_j} = \frac{\sum_{i=1}^{p+1} (I_i) - I_j}{I_j} = p \frac{1 - \delta_j/p}{1 + \delta_j} = p + \Delta p_j. \quad (\text{A.3})$$

The relative error in this ratio is then

$$\frac{\Delta p_j}{p} = -\frac{p+1}{p} \frac{\delta_j}{1 + \delta_j} \simeq -\frac{p+1}{p} \delta_j \quad (\delta_j \ll 1). \quad (\text{A.4})$$

This can be expressed as an error in $\Delta V_{BE,j}$:

$$\begin{aligned}\Delta V_{BE,j} - \Delta V_{BE}|_{\Delta p=0} &= \frac{kT}{q} \ln(p + \Delta p_j) - \frac{kT}{q} \ln(p) \\ &= \frac{kT}{q} \ln\left(1 + \frac{\Delta p_j}{p}\right) \simeq \frac{kT}{q} \frac{\Delta p_j}{p} = -\frac{kT}{q} \frac{p+1}{p} \delta_j.\end{aligned}\quad (\text{A.5})$$

A.1.2 With DEM

Dynamic element matching entails that each of the $p+1$ current sources is successively used as unit current source, and that the resulting ΔV_{BE} 's are then averaged. Thus, the first-order error terms will cancel. To estimate the residual second-order error in the average, we first rewrite the error (A.5) using the expansion $\ln(1 + \delta) = \delta - \delta^2/2 + O(\delta^3)$:

$$\Delta V_{BE,j} - \Delta V_{BE}|_{\Delta p=0} = \underbrace{\frac{kT}{q} \frac{\Delta p_j}{p}}_{\text{first-order error}} - \underbrace{\frac{kT}{2q} \left(\frac{\Delta p_j}{p}\right)^2}_{\text{second-order error}} + O(\Delta p^3).\quad (\text{A.6})$$

Neglecting the third- and higher-order terms, the average error is then

$$\begin{aligned}\Delta V_{BE,avg} - \Delta V_{BE}|_{\Delta p=0} &= \frac{1}{p+1} \frac{kT}{q} \sum_{j=1}^{p+1} \left(\frac{\Delta p_j}{p} - \frac{1}{2} \left(\frac{\Delta p_j}{p}\right)^2 \right) \\ &= \frac{1}{2(p+1)} \frac{kT}{q} \sum_{j=1}^{p+1} \left(\frac{\Delta p_j}{p}\right)^2.\end{aligned}\quad (\text{A.7})$$

If $|\Delta p_j/p| \leq \Delta p/p$, this average error is bounded as follows:

$$\left| \Delta V_{BE,avg} - \Delta V_{BE}|_{\Delta p=0} \right| < \frac{1}{2} \frac{kT}{q} \left(\frac{\Delta p}{p}\right)^2.\quad (\text{A.8})$$

Appendix B

Resolution Limits of Sigma-Delta Modulators with a DC Input

In this appendix, expressions are derived for the maximum resolution that can be obtained from a $\Sigma\Delta$ modulator operated as an incremental converter. First-order modulators and a single-loop second-order modulators are considered, both with ideal integrators, and with (more realistic) leaky integrators.

B.1 First-Order Modulator

B.1.1 Time-Domain Description

Consider the discrete-time model of a first-order $\Sigma\Delta$ modulator shown in Figure 4.8 of Section 4.3. Assuming that the input voltage and reference voltage are constant, $V_{IN}(n) = V_{IN}$ and $V_{REF}(n) = V_{REF}$, the output $V_{int1}(n)$ of the first integrator of this modulator at the *end* of the n^{th} clock cycle can be written as

$$V_{int1}(n) = V_{int1}(n-1) + a_1 \{V_{IN} - V_{REF} \cdot bs(n)\}, \quad (\text{B.1})$$

where a_1 is the gain of the integrator and $bs(n)$ is the value of the bitstream *during* the n^{th} clock cycle. The latter is determined by the comparator at the end of the previous clock cycle:

$$bs(n) = \begin{cases} 0 & \text{if } V_{int1}(n-1) \leq 0, \\ 1 & \text{if } V_{int1}(n-1) > 0. \end{cases} \quad (\text{B.2})$$

Incremental operation implies that the integrator is reset at the start of a conversion ($V_{int1}(0) = 0$), and hence $bs(1) = 0$.

B.1.2 Resolution Limit without Leakage

The maximum resolution of an incremental $\Sigma\Delta$ modulator that is operated during N cycles, is determined by the largest range of input levels that give rise to the same bitstream of length N . Such a range is sometimes referred to as a ‘dead zone’, as the output remains the same while the input signal changes within the range. For a modulator with ideal integrators, the widest dead zone can always be made more narrow by increasing the number of cycles N . Thus, the resolution can be increased indefinitely. This is not true for a modulator with leaky integrators, as will be explained in the next section.

The simulation result in Figure 4.10a shows that the widest dead zones of a first-order modulator occur at the extremes of the input range. This can also be derived analytically (see [1]). Consider an input voltage V_{IN} close to V_{REF} . This will cause the integrator's output to jump to $a_1 V_{IN}$ after the reset, i.e. $V_{int1}(1) = a_1 V_{IN}$. In the following clock cycles, the modulator will try to bring V_{int1} back to zero, producing a bitstream 01111... If $V_{IN} = V_{REF}$, the integrator's output will never reach zero, and the bitstream remains 1. For V_{IN} slightly smaller than V_{REF} , V_{int1} will reach zero after

$$N = \frac{V_{REF}}{V_{REF} - V_{IN}} \quad (\text{B.3})$$

clock cycles. Therefore, the width ΔV_{IN} of the dead zone that corresponds to a 01111... bitstream of length N equals V_{REF}/N . The maximum resolution that can be obtained from a first-order modulator is therefore

$$\text{ENOB}_{1st,ideal} = \log_2 \frac{V_{REF}}{\Delta V_{IN}} = \log_2 N. \quad (\text{B.4})$$

B.1.3 Resolution Limit with Leakage

Practical integrators are leaky as a result of their finite DC gain. As shown in Figure 4.12, this effect can be included in the model of the first-order modulator by adding a gain $p_1 < 1$ in the feedback of the integrator [2]. The leakage-free case corresponds to $p_1 = 1$.

The resolution of an incremental $\Sigma\Delta$ modulator with leaky integrators cannot be increased indefinitely by increasing the number of cycles N . This is because leakage causes the modulator to lock into a limit cycle, corresponding to a periodic bitstream that persists over a range of input values (a dead zone) [2].

The maximum ENOB that can be obtained for a given leakage is determined by the width ΔV_{IN} of the widest dead zone. The simulation results of Figure 4.13 indicate that this widest dead zone occurs around $V_{IN}/V_{REF} = 1/2$. This dead zone corresponds to a 01... limit cycle, which has a period of 2 cycles. The range of input values for which this limit cycle occurs can be found using Tsytkin's method [2]: assuming that the limit cycle occurs, the integrator outputs during the limit cycle are expressed as a function of V_{IN} ; then, conditions for V_{IN} can be found for which these outputs have polarities that agree with the comparator's output during the limit cycle.

A 01... limit cycle with a period of 2 cycles implies that the output of the integrator in any clock cycle has to be equal to the output two cycles later, $V_{int1}(n) = V_{int1}(n+2)$, and that the polarity of V_{int1} changes every clock cycle. Lets assume that $V_{int1}(n) \leq 0$, so that $bs(n+1) = 0$, and $V_{int1}(n+1) > 0$. Using the model of Figure 4.12, it can then be shown that

$$\begin{aligned} V_{int1}(n+2) &= p_1^2 V_{int1}(n) + a_1(1+p_1)V_{IN} - a_1 V_{REF} \\ &= V_{int1}(n). \end{aligned} \quad (\text{B.5})$$

Solving for $V_{int1}(n)$ gives

$$V_{int1}(n) = a_1 \frac{(1+p_1)V_{IN} - V_{REF}}{1-p_1^2}. \quad (\text{B.6})$$

An expression for $V_{int1}(n+1)$ in terms of V_{IN} and V_{REF} can be derived from this expression. Using these two expressions, the range of input values can be found for which $V_{int1}(n) \leq 0$ and $V_{int1}(n+1) > 0$:

$$\frac{p_1 V_{REF}}{1+p_1} < V_{IN} \leq \frac{V_{REF}}{1+p_1}. \quad (\text{B.7})$$

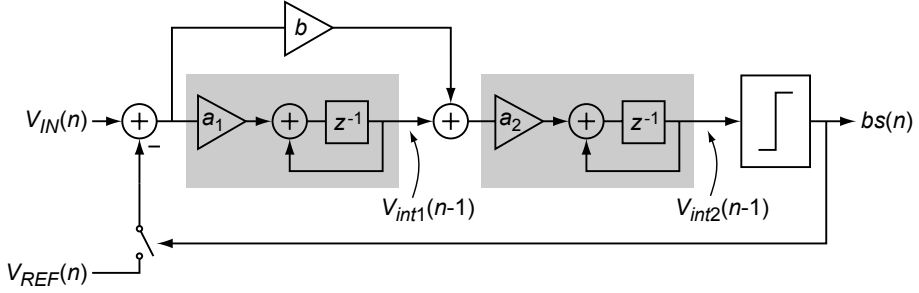


Figure B.1. Discrete-time model of a second-order $\Sigma\Delta$ modulator.

The width of this range, which is the width of the dead zone, equals

$$\Delta V_{IN} = \frac{1 - p_1}{1 + p_1} V_{REF}. \quad (\text{B.8})$$

The same result has been derived by Feely and Chua in [2]. Note that this width is zero if $p_1 = 1$, indicating that the 01... bitstream in the case of a leakage-free modulator is only produced by $V_{IN}/V_{REF} = 0.5$.

The effective number of bits obtained with an optimal decimation filter is therefore bounded as follows:

$$\text{ENOB}_{1st,leaky} \leq \log_2 \left(\frac{V_{REF}}{\Delta V_{IN}} \right) = \log_2 \left(\frac{1 + p_1}{1 - p_1} \right). \quad (\text{B.9})$$

B.2 Second-Order Single-Loop Modulator

B.2.1 Time-Domain Description

Consider the modulator of Figure B.1. The recursive expression for the output of the first integrator $V_{int1}(n)$ at the end of the n^{th} clock cycle is the same as for the first-order modulator, given by (B.1). The expression for the second integrator $V_{int2}(n)$ at the end of the n^{th} clock cycle is

$$\begin{aligned} V_{int2}(n) = & V_{int2}(n-1) + a_2 V_{int1}(n-1) + \\ & a_2 b \{ V_{IN} - V_{REF} \cdot bs(n) \}, \end{aligned} \quad (\text{B.10})$$

where $bs(n)$, as before, is the value of the bitstream *during* the n^{th} clock cycle. The comparator produces the bitstream based on the output of the second integrator:

$$bs(n) = \begin{cases} 0 & \text{if } V_{int2}(n-1) \leq 0, \\ 1 & \text{if } V_{int2}(n-1) > 0. \end{cases} \quad (\text{B.11})$$

Incremental operation means that both integrators are reset at the start of a conversion ($V_{int1}(0) = V_{int2}(0) = 0$), and hence $bs(1) = 0$.

B.2.2 Resolution Limit without Leakage

As for the first-order modulator, the maximum resolution that can be obtained from a second-order incremental $\Sigma\Delta$ modulator operated during N clock cycles is determined by the widest

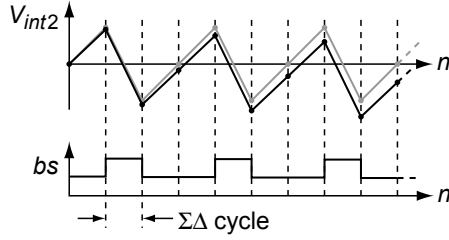


Figure B.2. Bitstream and output of the second integrator for $V_{IN}/V_{REF} = 1/3$ (gray) and for V_{IN}/V_{REF} slightly smaller than $1/3$ (black).

dead zone. Since the modulator becomes unstable near the extremes of the input range, a restricted input range will be considered. In that case, the widest dead zones correspond to the $010\dots$ and $011\dots$ limit cycles, which occur around $V_{IN}/V_{REF} = 1/3$ and $2/3$, respectively. ($V_{IN}/V_{REF} = 1/2$ leads to a $0011\dots$ limit cycle which corresponds to a narrower dead zone).

Consider the case $V_{IN}/V_{REF} = 1/3$. Using the time-domain description of the modulator, it can be shown that the output of the second integrator periodically takes the values 0 , $a_2 b V_{REF}/3$, and $-a_2(b - a_1)V_{REF}/3$, leading to the bitstream $010\dots$. If V_{IN}/V_{REF} is slightly larger than $1/3$, this pattern is immediately disturbed: the bitstream then starts with $0101\dots$. If, in contrast, V_{IN}/V_{REF} is slightly smaller than $1/3$, the $010\dots$ pattern is maintained for a certain number of clock cycles. The output of the second integrator slowly drifts down, until its positive peak reaches zero (see Figure B.2). At that point, the $010\dots$ pattern is broken.

The width of the dead zone around $1/3$ corresponds to the maximum deviation ΔV_{IN} from $V_{REF}/3$ for which the modulator still produces a $010\dots$ bitstream of length N . The drift ΔV_{int2} of the output of the second integrator can be found by calculating the response of the loop filter to a constant input ΔV_{IN} . After N clock cycles, the integrator's output will have drifted down by

$$\Delta V_{int2}(N) = \Delta V_{IN} a_2 \left(bN + \frac{1}{2} a_1 N(N-1) \right). \quad (\text{B.12})$$

By equating this to the peak value $a_2 b V_{REF}/3$ and solving for ΔV_{IN} , the width of the dead zone is found:

$$\Delta V_{IN} = \frac{b/3}{bN + \frac{1}{2} a_1 N(N-1)} V_{REF}. \quad (\text{B.13})$$

The maximum resolution that can be obtained from a second-order modulator is therefore

$$\text{ENOB}_{2nd,ideal} = \log_2 \frac{V_{REF}}{\Delta V_{IN}} = \log_2 \left(3N + \frac{3a_1}{2b} N(N-1) \right) \quad (\text{B.14})$$

$$\simeq 2 \log_2(N) - \log_2 \left(\frac{2b}{3a_1} \right). \quad (\text{B.15})$$

B.2.3 Resolution Limit with Leakage

As for the first-order modulator, leakage causes a second-order modulator to lock into limit cycles, resulting in dead zones that limit the achievable resolution irrespective of the number of cycles. Leakage can be included in the model of a second-order modulator (Figure B.1) by using non-unity feedback coefficients p_1 and p_2 in the first and second integrator, respectively.

If a limited input range is considered (avoiding the unstable regions near zero and V_{REF}), the widest dead zone is again found around $V_{IN}/V_{REF} = 1/3$, and corresponds to the 010... limit cycle. As for the first-order modulator, the width of this dead zone can again be found using Tsytkin's method. From the equations $V_{int1}(n) = V_{int1}(n+3)$ and $V_{int2}(n) = V_{int2}(n+3)$, expressions for the periodic outputs of both integrators in terms of V_{IN} and V_{REF} can be derived. The range of input values can then be found for which the polarity of the output of the second integrator agrees with the 010... bitstream. After some manipulation, the width ΔV_{IN} of this range is found as:

$$\Delta V_{IN} = \frac{b(1-p_1^3)(p_2-p_2^2) + a_1(1-p_1-p_2+p_1p_2^2+p_1^2p_2-p_1^2p_2^2)}{(1+p_2+p_2^2)(b(1-p_1^3)+a_1(1+p_1+p_1^2))} V_{REF}, \quad (\text{B.16})$$

which can be simplified to

$$\Delta V_{IN} \simeq (1-p_1)(1-p_2) \frac{b}{3a_1} V_{REF} \quad (\text{B.17})$$

for p_1 and p_2 close to 1. The maximum resolution that can be obtained is therefore

$$\text{ENOB}_{2nd,leaky} \leq \log_2 \left(\frac{V_{REF}}{\Delta V_{IN}} \right) \simeq \log_2 \left(\frac{3a_1}{b(1-p_1)(1-p_2)} \right) \quad (\text{B.18})$$

$$= \log_2 \left(\frac{3A_{0,1}A_{0,2}}{ba_2} \right) \quad (A_{0,1} \gg a_1, A_{0,2} \gg a_2). \quad (\text{B.19})$$

References

- [1] J. Robert and P. Deval, "A second-order high-resolution incremental A/D converter with offset and charge injection compensation," *IEEE Journal of Solid-State Circuits*, vol. 23, no. 3, pp. 736–741, June 1988.
- [2] O. Feely and L. O. Chua, "The effect of integrator leak in Σ - Δ modulation," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 11, pp. 1293–1305, Nov. 1991.

Appendix C

Non-Exponential Settling Transients

When the base-emitter voltage of a diode-connected bipolar transistor is sampled on a capacitor, the settling transient will be non-exponential, as a result of the non-linear voltage-current characteristic of the transistor. In this appendix, an expression will be derived for the time required for the voltage on the sampling capacitor to settle with a given accuracy to its final value.

C.1 Problem Description

A circuit model of the non-linear settling problem is shown in Figure C.1. A diode-connected bipolar transistor is connected to a capacitor C (in a switched-capacitor circuit this would be the sampling capacitor) via a resistor R (modeling, for instance, the on-resistance of a switch). Initially, the transistor is biased at a current I_1 , resulting in a base-emitter voltage V_{BE1} , and the voltage V_C across the capacitor has settled to V_{BE1} . At time $t = 0$, the bias current changes to I_2 . As a result V_C will eventually change to V_{BE2} . Since the circuit is non-linear, this transient will be non-exponential, because the impedance of the transistor will change during the transient.

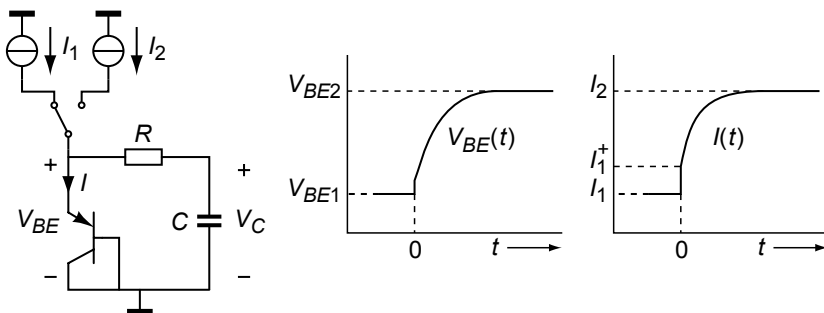


Figure C.1. Circuit model of the non-exponential settling problem.

C.2 Settling Transients from $V_{BE1} \neq 0$ to V_{BE2}

If, at a given time $t > 0$ during the transient, the current through the transistor is $I(t)$, the current flowing into the capacitor is $I_2 - I(t)$. As a result, the voltage across the capacitor will be changing with a rate

$$\frac{dV_C}{dt} = \frac{I_2 - I}{C}. \quad (\text{C.1})$$

At this operating point, the small-signal transconductance of the transistor is

$$g_m = \frac{qI}{kT}. \quad (\text{C.2})$$

The rate of change of I can then be derived as

$$\frac{dI}{dt} = \frac{1}{1/g_m + R} \frac{dV_C}{dt} = \frac{I_2 - I}{\left(\frac{kT}{qI} + R\right)C}. \quad (\text{C.3})$$

A solution to this non-linear differential equation, which would describe I as a function of t , is hard to find analytically.

Fortunately, it is possible to find an implicit solution that expresses t as a function of I . Thus, given a current I , the time instant at which this current occurs can be calculated. To find this implicit solution, the differential equation is rewritten as

$$dt = \frac{\left(\frac{kT}{qI} + R\right)C}{I_2 - I} dI = \left(\frac{kTC}{qI(I_2 - I)} + \frac{RC}{I_2 - I}\right) dI. \quad (\text{C.4})$$

Integration gives

$$t = -\frac{kTC}{qI_2} C \ln\left(\frac{I_2 - I}{I}\right) - RC \ln(I_2 - I) + K, \quad (\text{C.5})$$

where it is assumed that $I_2 > I$ (i.e. it is a rising transient).

To determine the constant K , the initial conditions at time $t = 0^+$ have to be substituted, where $t = 0^+$ refers to the moment just after the bias current has changed to I_2 . At this moment, the current through the bipolar transistor is $I(0^+) = I_1^+$, which is not equal to I_1 unless $R = 0$. Its exact value can be derived from

$$V_{BE}(0^+) = \frac{kT}{q} \ln\left(\frac{I_1^+}{I_S}\right) = V_C(0^+) + R(I_2 - I_1^+), \quad (\text{C.6})$$

$$V_C(0^+) = V_C(0) = V_{BE1} = \frac{kT}{q} \ln\left(\frac{I_1}{I_S}\right), \quad (\text{C.7})$$

which leads to the equation

$$R(I_2 - I_1^+) = \frac{kT}{q} \ln\left(\frac{I_1^+}{I_1}\right), \quad (\text{C.8})$$

which can be solved numerically for I_1^+ . The constant K can now be found by substituting $t = 0$, $I = I_1^+$ in (C.5), which leads to

$$t = \frac{kTC}{qI_2} C \ln\left(\frac{(I_2 - I_1^+)I}{(I_2 - I)I_1^+}\right) + RC \ln\left(\frac{I_2 - I_1^+}{I_2 - I}\right), \quad (\text{C.9})$$

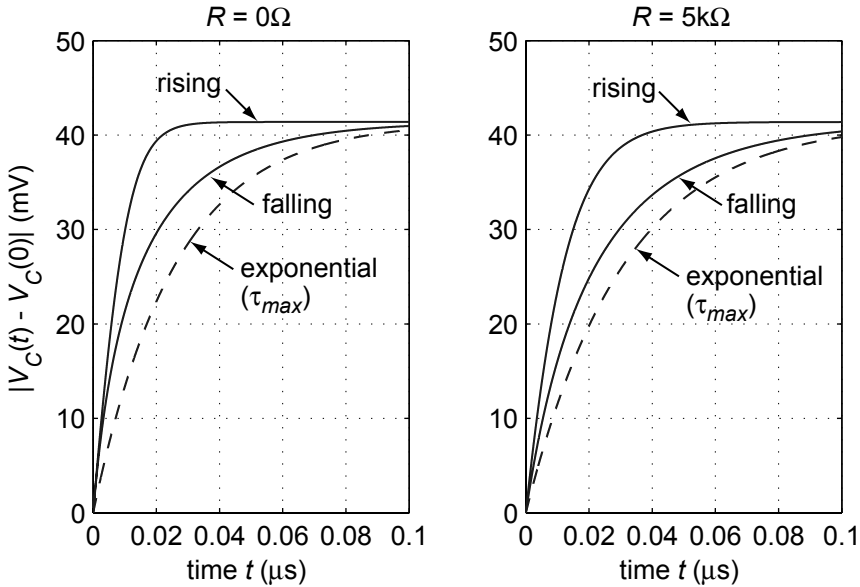


Figure C.2. Rising and falling step-responses of V_C when the bias current is switched between $1 \mu\text{A}$ and $5 \mu\text{A}$, for $C = 1 \text{ pF}$, $T = 300 \text{ K}$ and $R = 0 \Omega, 5 \text{ k}\Omega$. The dashed lines are exponential approximations. (The falling step responses have been inverted for easier comparison.)

which is valid for both rising and falling transients. This expression shows that the settling is only exponential (with time constant $\tau = RC$) if $R \gg kT/qI_2$ (which is usually not the case). For $R = 0$, the expression simplifies to

$$t = \frac{kT}{qI_2} C \ln \left(\frac{(I_2 - I_1) I}{(I_2 - I) I_1} \right) \quad (R = 0). \quad (\text{C.10})$$

Figure C.2 shows step responses that correspond to equation (C.9), found by calculating the values of V_C and t that correspond to currents I in the range of $1 \mu\text{A}$ to $5 \mu\text{A}$ and a capacitance C of 1 pF . The falling step responses ($I_1 > I_2$) have been inverted for easier comparison. Note the asymmetry between the rising and falling transients. If $R = 0$, the initial slope of the rising and falling transient is the same and equals $(I_2 - I_1)/C = 4 \text{ V}/\mu\text{s}$. The figure also shows exponential settling based on the worst-case time constant τ_{max} of the circuit, which is achieved at the smallest bias current I_{min} :

$$\tau_{max} = \left(\frac{kT}{qI_{min}} + R \right) C. \quad (\text{C.11})$$

For rising transients, $I_{min} = I_1$, while for falling transients, $I_{min} = I_2$. The figure shows that the circuit always settles faster than this exponential.

To calculate the time t_{settle} at which V_C is a fraction ε of the step ΔV_{BE} away from its final value, the corresponding current I_{settle} has to be found from

$$V_C(t_{settle}) = \frac{kT}{q} \ln \frac{I_{settle}}{I_S} - R(I_2 - I_{settle}) \quad (C.12)$$

$$= V_{BE2} - \varepsilon \Delta V_{BE} = \frac{kT}{q} \left\{ \ln \frac{I_2}{I_S} - \varepsilon \ln \frac{I_2}{I_1} \right\} \quad (C.13)$$

$$\Rightarrow R(I_2 - I_{settle}) = \frac{kT}{q} \left\{ \ln \frac{I_{settle}}{I_2} + \varepsilon \ln \frac{I_2}{I_1} \right\}. \quad (C.14)$$

The current I_{settle} can be solved numerically from this equation. If the voltage drop across the resistor R is negligible (a requirement that is met for small values of ε and R and that can be checked after calculating I_{settle}), an analytical solution can be found:

$$I_{settle} = I_2 \left(\frac{I_2}{I_1} \right)^{-\varepsilon}. \quad (C.15)$$

Substitution of this value in (C.9) yields:

$$t_{settle} = \frac{kT}{qI_2} C \{ \ln p^+ - \varepsilon \ln p \} + \left(R + \frac{kT}{qI_2} \right) C \ln \left(\frac{p^+ - 1}{p^+ (1 - p^{-\varepsilon})} \right), \quad (C.16)$$

where $p = I_2/I_1$ and $p^+ = I_2/I_1^+$. For $R = 0$, the expression can again be simplified to

$$t_{settle} = \frac{kT}{qI_2} C \left\{ \ln \left(\frac{p - 1}{1 - p^{-\varepsilon}} \right) - \varepsilon \ln p \right\} \quad (R = 0). \quad (C.17)$$

Figure C.3 shows the settling times predicted by these equations, for various values of the current ratio p . The times shown in this figure are normalized to a time constant τ given by

$$\tau = \frac{kT}{qI_{min}}, \quad (C.18)$$

where I_{min} is, as before, the smallest bias current, which is I_1 for rising transients, and I_2 for falling transients. The figure shows the cases $R = 0$ (left) and $R = \tau/2C$ (right). It also shows that the settling time is bounded by that required in the case of pure exponential settling with a time constant τ_{max} given by equation (C.11).

C.3 Settling Transients from $V_{BE1} = 0$ to V_{BE2}

To include the situation $I_1 = 0$ in the analysis, where the capacitor is initially completely discharged and the base-emitter voltage is zero, the transistor's saturation current I_S cannot be ignored. The initial current I_1^+ can then be derived from

$$R(I_2 - I_1^+) = \frac{kT}{q} \ln \left(\frac{I_1^+ + I_S}{I_S} \right), \quad (C.19)$$

Since I_1^+ will be close to $I_1 = 0$, it can be approximated by

$$I_1^+ = I_S \left\{ \exp \left(\frac{qRI_2}{kT} \right) - 1 \right\}. \quad (C.20)$$

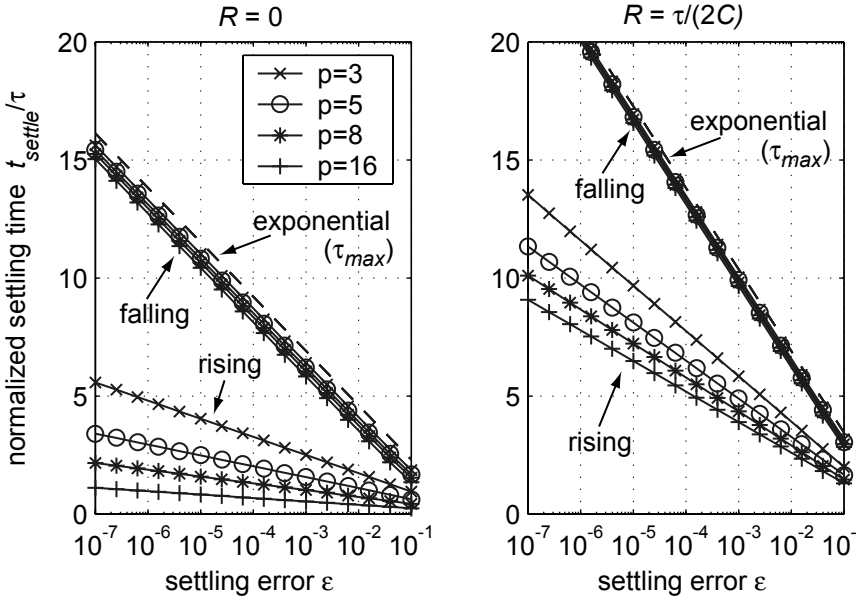


Figure C.3. Settling time required for V_C to settle to $\epsilon \cdot \Delta V_{BE}$ from its final value, normalized to the time constant τ given by (C.18), for both rising and falling transients (solid lines), and for the worst-case exponential approximation with time constant given by (C.11) (dashed lines).

The transconductance has to be replaced by

$$g_m = \frac{q(I + I_S)}{kT}. \quad (C.21)$$

The differential equation then becomes

$$dt = \frac{\left(\frac{kT}{q(I+I_S)} + R\right) C}{I_2 - I} dI = \left(\frac{kTC}{q(I + I_S)(I_2 - I)} + \frac{RC}{I_2 - I}\right) dI, \quad (C.22)$$

and its solution

$$t = \frac{kT}{q(I_2 + I_S)} C \ln \left(\frac{(I_2 - I_1^+)(I + I_S)}{(I_2 - I)(I_1^+ + I_S)} \right) + RC \ln \left(\frac{I_2 - I_1^+}{I_2 - I} \right). \quad (C.23)$$

Since usually $I_2, I \gg I_S$, this can be simplified to

$$t = \frac{kT}{qI_2} C \ln \left(\frac{(I_2 - I_1^+) I}{(I_2 - I)(I_1^+ + I_S)} \right) + RC \ln \left(\frac{I_2 - I_1^+}{I_2 - I} \right). \quad (C.24)$$

Substituting (C.20) and using $I_1^+ \ll I_2$ gives

$$t = \frac{kT}{qI_2} C \ln \left(\frac{I_2 \cdot I}{(I_2 - I) I_S} \right) + RC \left\{ \ln \left(\frac{I_2}{I_2 - I} \right) - 1 \right\}. \quad (C.25)$$

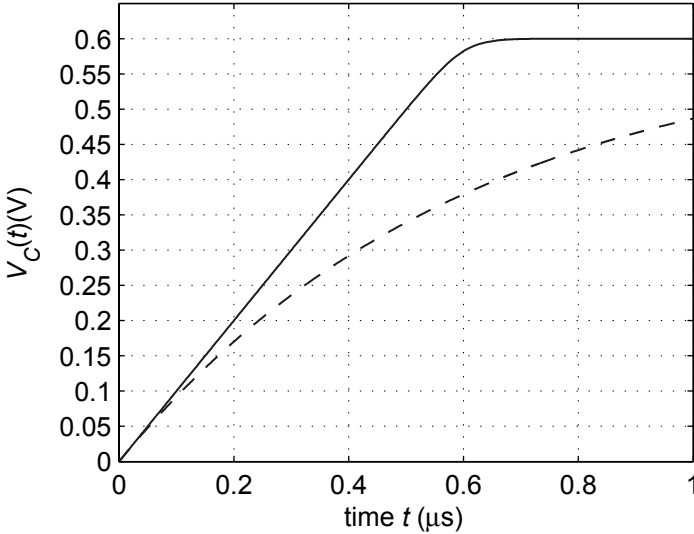


Figure C.4. Transient step-response of V_C when the bias current is switched from $I_1 = 0$ to $I_2 = 1 \mu\text{A}$, with $R = 0$, $C = 1 \text{ pF}$, $T = 300 \text{ K}$, and $I_S \simeq 4 \cdot 10^{-17} \text{ A}$. The dashed line is an exponential step-response with the same initial slope.

For $R = 0$, $I_1^+ = I_1 = 0$, the expression can be further simplified to

$$t = \frac{kT}{qI_2} C \ln \left(\frac{I_2 \cdot I}{(I_2 - I) I_S} \right) \quad (R = 0). \quad (\text{C.26})$$

Figure C.4 shows a transient response that corresponds to this equation, found by calculating the values of V_C and t that correspond to currents in the range 0 to $I_2 = 1 \mu\text{A}$ and a capacitance C of 1 pF. Note the clear slewing behaviour during the initial $0.5 \mu\text{s}$, which is very dissimilar from exponential settling. During this time, virtually all current flows into C , giving a slew-rate of $I_2/C = 1 \text{ V} / \mu\text{s}$.

As in the previous section, to calculate the time t_{settle} at which V_{BE} is a fraction ε below its final value, the corresponding current I_{settle} has to be found from

$$V_{BE}(t_{\text{settle}}) = \frac{kT}{q} \ln \frac{I_{\text{settle}}}{I_S} - R(I_2 - I_{\text{settle}}) \quad (\text{C.27})$$

$$= (1 - \varepsilon)V_{BE2} = (1 - \varepsilon) \frac{kT}{q} \ln \frac{I_2}{I_S} \quad (\text{C.28})$$

$$\Rightarrow R(I_2 - I_{\text{settle}}) = \frac{kT}{q} \left\{ \ln \frac{I_{\text{settle}}}{I_2} - \varepsilon \ln \frac{I_2}{I_S} \right\}. \quad (\text{C.29})$$

This again only gives I_{settle} indirectly. If the voltage drop across the resistor R is negligible, an analytical solution can be found:

$$I_{\text{settle}} = I_S \left(\frac{I_2}{I_S} \right)^{1-\varepsilon}. \quad (\text{C.30})$$

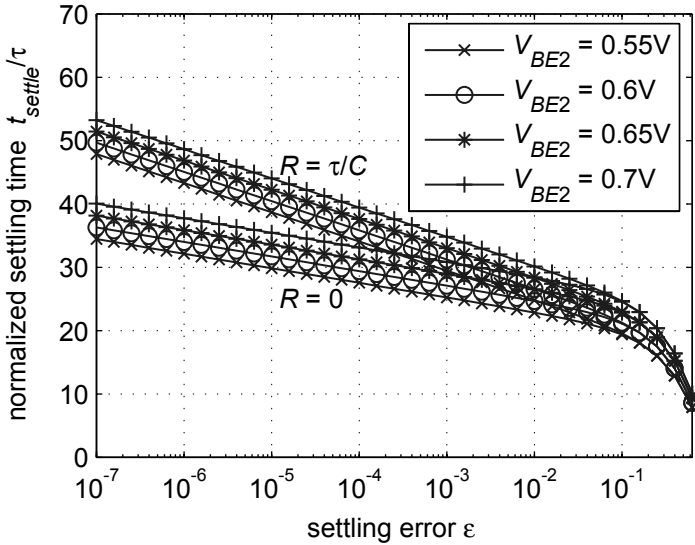


Figure C.5. Time required for V_C to settle to $(1 - \varepsilon) \cdot V_{BE2}$, normalized to the time constant τ of the final exponential settling if $R = 0$, for $R = 0$ and $R = C/\tau$, and for various values of V_{BE2} .

This value can then be substituted for I in (C.24) or (C.26). For the former, the result can be rewritten as

$$t_{settle} = \frac{kT}{qI_2} C \left\{ \ln \left(\frac{I_S}{(I_1^+ + I_S)} \right) + (1 - \varepsilon) \ln \left(\frac{I_2}{I_S} \right) \right\} + \left(R + \frac{kT}{qI_2} \right) C \left\{ \ln \left(\frac{I_2 - I_1^+}{I_2} \right) - \ln \left(1 - \left(\frac{I_2}{I_S} \right)^{-\varepsilon} \right) \right\}, \quad (C.31)$$

while for the latter this can be simplified to

$$t_{settle} = \frac{kT}{qI_2} C \left\{ (1 - \varepsilon) \ln \left(\frac{I_2}{I_S} \right) - \ln \left(1 - \left(\frac{I_2}{I_S} \right)^{-\varepsilon} \right) \right\} \quad (R = 0). \quad (C.32)$$

Figure C.5 shows the settling time predicted by these equations as a function of the final base-emitter voltage V_{BE2} (which corresponds to a certain I_2/I_S). The settling time in the figure is normalized to

$$\tau = \frac{kTC}{qI_2}, \quad (C.33)$$

which is the time constant of the final exponential settling if $R = 0$. For large of ε , which correspond to the slewing part of the step response, the settling time is a linear function of ε (first term in (C.32)). For smaller values of ε , the settling is exponential (second term in (C.32)). The transition between slewing and exponential settling occurs at roughly

$$t_{slew_end} = \tau \ln \left(\frac{I_2}{I_S} \right) = \tau \frac{qV_{BE2}}{kT}. \quad (C.34)$$

A larger value of V_{BE2} implies a longer slewing period and therefore a positive offset on the settling time. A non-zero resistance value has little effect on the slewing phase but changes the time constant of the final exponential settling to

$$\tau' = \left(R + \frac{kTC}{qI_2} \right) C. \quad (\text{C.35})$$

SUMMARY

This book describes the analysis and design of precision temperature sensors in CMOS technology. It focuses on so-called ‘smart’ temperature sensors, which provide a readily interpretable digital output. Over the military temperature range, which extends from $-$ to 125°C , the inaccuracy obtained using such sensors in previous work was about $\pm 2^{\circ}\text{C}$. In this work, that inaccuracy is reduced to $\pm 0.1^{\circ}\text{C}$, so that the performance of CMOS temperature sensors becomes comparable to that of conventional sensors, such as platinum resistors and thermistors. To keep production costs low, a standard CMOS process is used, and the sensors are calibrated at only one temperature.

In a smart temperature sensor, an analog-to-digital converter (ADC) determines the ratio between a temperature-dependent voltage and a reference voltage. These voltages can be accurately generated using bipolar transistors. A voltage that is proportional to absolute temperature (PTAT) can be generated as the difference in base-emitter voltage ΔV_{BE} between two matched transistors operated at different current densities. A reference voltage can be obtained by combining a base-emitter voltage V_{BE} and a scaled ΔV_{BE} . The scale factor α is chosen such that the positive temperature coefficient of $\alpha \Delta V_{BE}$ compensates for the negative temperature coefficient of V_{BE} .

In CMOS technology, two types of bipolar transistors are available: substrate pnp transistors and lateral pnp transistors. Substrate pnp transistors are the device of choice for the implementation of smart temperature sensors, because their characteristics more closely resemble those of an ideal bipolar transistor. A disadvantage of these transistors is their grounded collector, which implies that they have to be biased via their emitter. As a result, their base-emitter voltage is not only a function of the saturation current and the bias current, but also of the current gain.

The accuracy of a smart temperature sensor depends on the accuracy of the voltages V_{BE} and ΔV_{BE} . The various non-idealities that affect V_{BE} and

ΔV_{BE} are discussed in detail. ΔV_{BE} can be made very accurate by design, provided that dynamic element matching (DEM) is applied to generate an accuracy current-density ratio. The main source of inaccuracy then lies in V_{BE} , which spreads due to processing variations and due to mechanical stress induced by the sensor's package. Unless this spread is trimmed out, the inaccuracy is limited to about $\pm 1^\circ\text{C}$. Spread of the nominal value of the transistor's saturation current and its bias current give rise to a PTAT spread of V_{BE} . Since such spread essentially introduces only one degree of freedom, it can be trimmed based on a calibration at a single temperature.

To prevent a non-PTAT spread of V_{BE} , which would require more calibration temperatures, the bias current can best be generated from a PTAT voltage using a resistor with a reproducible temperature coefficient. Current-gain spread leads to non-PTAT spread as a result of the current-gain dependency of V_{BE} . A special bias circuit is proposed that makes V_{BE} independent of the current gain. Packaging-induced stress also leads to non-PTAT spread. Since substrate pnp transistor are relatively insensitive to the tensile stress introduced by ceramic and metal-can packages, it is best to use these packages for precision sensors.

Trimming of V_{BE} can be implemented using various techniques. A modulated trimming technique is proposed, which adjusts the transistor's average bias current by switching its bias current back and forth between two values that correspond to the extremes of the trimming range. Using this technique, a high trimming resolution can be obtained, without the complexity and large chip area associated with conventional trimming techniques.

Before trimming, a calibration is needed to establish the initial error. This calibration is performed after packaging to include errors due to packaging stress. Conventional calibration techniques are then time-consuming and therefore costly. Three alternative low-cost calibration techniques are described: batch calibration, calibration based on ΔV_{BE} measurement, and voltage reference calibration. Batch calibration exploits the fact that the spread between sensors from one production batch is typically much smaller than batch-to-batch spread. The other two techniques are based on voltage measurements, which can be performed much faster than temperature measurements.

The base-emitter V_{BE} not only suffers from spread, but also from systematic non-linearity, or curvature. This can be removed using the various curvature-correction techniques developed for bandgap voltage references. A simpler solution is to make use of the ratiometric nature of smart temperature sensors, and to introduce a compensating non-linearity by using a slightly temperature-dependent reference voltage. An extension of this technique for higher-order curvature correction is presented.

The ADC in a smart temperature sensor typically has to produce only about 10 readings per second, because the signal bandwidth is limited by the thermal properties of the package. For precision sensors, a resolution in the order of

15 bits ($0.01\text{ }^{\circ}\text{C}$) is desired. Indirect ADCs, and specifically those based on a sigma-delta modulator, are the best choice for these requirements. The use of a second-order sigma-delta modulator allows the requirements to be met at a modest clock frequency of about 10 kHz, which reduces power consumption and errors related to switching transients compared to a first-order modulator. A simple charge-balancing scheme, in which V_{BE} and ΔV_{BE} are alternately integrated, can be used to obtain the desired transfer function.

The output of a sigma-delta modulator is a bitstream, which has to be processed by a decimation filter to obtain a final conversion result. A filter with a simple triangular impulse response is a good compromise between resolution and complexity. A technique for fine-tuning the gain of this filter is proposed, so that the conversion result can be directly scaled to degrees Celsius. In addition, a linearization technique is introduced that allows an arbitrary non-linearity to be corrected in the decimation filter. This technique is based on a lookup table and requires less circuitry than comparable existing techniques. Finally, it is shown how the timing of dynamic error correction techniques (such as DEM and modulated trimming) has to be organized so that the related dynamic error signals are filtered out by the ADC. Specifically, it is shown how bitstream-controlled timing can be used to prevent errors due to intermodulation of the bitstream and dynamic error signals.

The implementation of CMOS smart temperature sensors using both continuous-time (CT) and switched-capacitor (SC) techniques is discussed. In both cases, the first integrator of the sigma-delta modulator must have an offset in the μV range, which requires advanced offset-cancellation techniques. For a CT implementation, the nested-chopper technique is an attractive technique, while for a SC implementation a combination of autozeroing and low-frequency chopping can be used. DEM can be used to eliminate errors due to component mismatch. A SC implementation is more compatible with DEM and other dynamic error correction techniques than a CT implementation, because it is less sensitive to switching transients and the finite on-resistance of switches. For the same noise performance, a CT implementation is more power-efficient. Fortunately, even for low noise levels, the power consumption of a SC implementation is not so high that self-heating would limit the accuracy.

Two CT realizations and one SC realization are presented, in which the techniques described in this book have been applied. In the latter, the most dynamic error correction techniques have been applied. It has been implemented in a $0.7\text{ }\mu\text{m}$ CMOS process, and has a 3σ inaccuracy of $\pm 0.1\text{ }^{\circ}\text{C}$ over the full military temperature range. A comparison with previous work shows that this is the highest reported accuracy to date.

About the Authors

Michiel A. P. Pertijs was born in Roosendaal, The Netherlands, on May 31, 1977. He received the M.Sc. and Ph.D. degrees in electrical engineering (both cum laude) from Delft University Technology in 2000 and 2005, respectively. Since August 2005, he is working as a circuit design engineer for National Semiconductor in Delft, The Netherlands.

From 2000 to 2005, he worked as a research assistant at the Electronic Instrumentation Laboratory of Delft University of Technology, on the subject of high-accuracy CMOS smart temperature sensors. In co-operation with Philips Semiconductors, his research has been applied in commercial temperature sensors, and has resulted in four patent applications. At Delft University, he has been involved in various teaching activities. Since 2003, he has been a lecturer in the Europractice Course on Smart Sensor Systems.

From 1997 to 1999, he worked for EARS B.V., Delft, on the design and production of a handheld photo-synthesis meter. In 2000, he was an intern with Philips Semiconductors, Sunnyvale, California, where he worked on the design of interface electronics for smart temperature sensors.

Dr. Pertijs received the ISSCC 2005 Jack Kilby Award for Outstanding Student Paper. His research interests include analog and mixed-signal electronics and smart sensors.

Johan H. Huijsing was born on May 21, 1938. He received the M.Sc. degree in Electrical Engineering from the Delft University of Technology, Delft, the Netherlands in 1969, and the Ph.D. degree from this University in 1981 for his thesis on operational amplifiers.

He has been an assistant and associate professor in Electronic Instrumentation at the Faculty of Electrical Engineering of the Delft University of Technology since 1969, where he became a full professor in the chair of Electronic Instrumentation since 1990, and professor-emeritus since 2003. From 1982 through 1983 he was a senior scientist at Philips Research Labs. in Sunnyvale,

California, USA. From 1983 until 2005 he was a consultant for Philips Semiconductors, Sunnyvale, California, USA and since 1998 also a consultant for Maxim, Sunnyvale, California, USA.

The research work of Johan Huijsing is focussed on the systematic analysis and design of operational amplifiers, analog-to-digital converters and integrated smart sensors. He is author or co-author of some 250 scientific papers, 40 patents and 13 books, and co-editor of 13 books. He is fellow of IEEE for contributions to the design and analysis of analog integrated circuits. He was awarded the title of Simon Stevin Meester for applied Research by the Dutch Technology Foundation.

He is initiator and co-chairman until 2005 of the international Workshop on Advances in Analog Circuit Design, which has been held annually, since 1992 in Europe. He has been a member of the programme committee of the European Solid-State Circuits Conference from 1992 until 2002. He has been chairman of the Dutch STW Platform on Sensor Technology and chairman of the biennial national Workshop on Sensor Technology from 1991 until 2002.

Index

- 1/ f noise
 - corner frequency, 179, 192, 200
 - dynamic offset cancellation, 176
 - of lateral bipolar transistors, 25
- ADC, 4, 6, 53, 69, 73, 88, 90, 107, 142, 221–222, 272
 - direct ADCs, 110
 - dual slope, 110, 112, 114–115
 - dynamic range, 54
 - indirect, 110
 - oversampling, 73
 - requirements, 107
 - See also* sigma-delta ADCs
- Asynchronous modulators, 112
 - comparison to synchronous, 113
- Auto-calibration, 214
- Autozeroing, 175
- Bandgap energy, 52
- Bandgap reference, 5, 33, 51, 78, 221, 228, 230–231
 - curvature correction, 80
 - curvature, 79
- Bandgap voltage, 3–5, 11
 - temperature dependency, 21
- Bandwidth, 110, 115–116, 119
- Base current, 18
 - compensation for, 232
 - components, 19
 - temperature dependency, 24
- Base doping, 19, 29, 31, 68
- Base transport factor, 19
- Batch calibration, 218, 228, 236, 258, 264
- Bias circuit, 6, 63
 - current-gain-dependent, 92
 - modified PTAT/R, 93, 250
 - PTAT/R, 66, 234
 - startup, 64
 - structure, 63
- Bias current, 64
 - CTAT/R, 65
 - PTAT/R, 65
 - TI/R, 66
- BiCMOS technology, 84
- Binning, 214
- Bipolar technology, 5, 82, 84
- Bipolar transistors
 - lateral, 5, 24, 27–28
 - vertical, 5, 24, 30, 34–35
- Bitstream, 113, 117
- Bus interface, 223, 229, 235, 241
- Calibration transistor, 219–220, 236
- Calibration, 2, 6, 31, 43, 45, 55, 69, 75, 87, 92, 213
 - after packaging, 7, 217
 - based on ΔV_{BE} measurement, 219, 258, 265
 - calibration report, 214
 - conventional techniques, 217, 263
 - extrapolation from, 215
 - ISO definition, 214
 - of smart sensors, 214
 - wafer-level, 7, 217
 - See also* auto-calibration, batch calibration, voltage reference calibration
- Capacitor
 - MOS, 185
- Capacitors, 185
 - double-poly, 185, 239
 - matching, 130
 - metal, 185
 - MOS, 185, 229, 239
 - non-linearity, 185, 239
- Cascoded current sources, 63
- Charge balancing, 111
 - implementation, 159, 229, 246
 - in asynchronous ADCs, 111
 - in CT circuitry, 165
 - in SC circuitry, 182
 - in synchronous ADCs, 113
 - in temperature sensors, 111

- Charge injection, 193, 196, 201–202
- Chopper amplifier, 176, 235
 - delayed demodulation, 199
 - filtering of spike harmonics, 199
 - guard time, 199
 - residual offset, 178, 196
 - spike suppression, 199
 - See also* nested-chopper technique
- Chopping, 168, 172, 176–177
 - bitstream-controlled, 150
 - in bias circuit, 251
 - of a sigma-delta modulator, 255
 - of an autozeroed amplifier, 202
 - of an autozeroed integrator, 203
 - of switches, 202
 - system-level, 206
- Clipping, 121, 131–132
- Clock boosting, 255
- Clock feed-through, 178, 193
- Clock jitter, 168, 174
- CMOS technology, 2, 4–5, 18, 24, 26–28, 30, 36, 40, 45, 51, 84, 206
- Collector current, 16–17
 - dependency on base-collector voltage, 37
 - of a diode-connected transistor, 18, 23
 - of lateral transistors, 26
 - temperature dependency, 23
- Common-base current-gain, 19–20, 31
- Common-centroid layout, 34, 59, 168, 184, 195, 235
- Common-emitter current-gain, 19, 23, 27, 31
- Comparator, 111–112, 117
 - implementation, 255
- Continuous-time circuitry, 165
- Continuous-time integrator, 160
- Continuous-time loop filter, 117
- Current crowding, 26, 98
- Current gain
 - compensation for finite, 92, 250
 - current dependency, 20, 27
 - effect of mechanical stress, 35
 - errors due to finite, 92
 - processing spread, 31
 - temperature dependency, 24
 - See also* common-emitter current gain, common-base current gain
- Curvature correction, 78
 - classification, 80
 - comparison, 87
 - digital, 91
 - in decimation filter, 143, 248
 - miscellaneous techniques, 86
 - piecewise linear, 86
 - ratiometric, 89
 - See also* ratiometric curvature correction
 - system level, 91
 - temperature-dependent bias current ratio, 82
 - temperature-dependent bias current, 81
 - temperature-dependent gain, 82
- Curvature, 22, 53, 55, 78
 - due to bias resistor, 42
 - due to capacitor non-linearity, 185
 - due to finite current-gain, 24, 92
 - due to resistor non-linearity, 169
 - due to saturation current, 22
 - due to stress, 34
 - in voltage references, 79
 - temperature errors due to, 78
- Data-weighted averaging, 150
- DC gain, 125, 128, 136–137, 170, 177, 187, 192, 252, 278
- Decimation filter, 117
 - based on window function, 139
 - curvature correction, 143
 - in incremental ADC, 120
 - linear scaling, 142
 - matched to loop filter, 138
 - optimal, 122, 134
 - $\text{sinc}^2/\text{triangular}$, 123, 141
 - $\text{sinc}^3/\text{quadratic}$, 140–141
 - $\text{sinc}/\text{rectangular}$, 123, 140–141
 - timing, 257
 - zeros, 147
- Differential circuitry, 179, 198
 - charge injection in, 197
 - leakage in, 186
 - linearity, 185
- Diffusion current, 12, 15–17, 19, 25
- Digital interface, 4–5, 114, 220, 223
- Diode, 11, 272
 - ideal characteristic, 12
 - one-sided, 15
 - recombination, 13
- Dither, 124
- Dummy switches, 196
- Dynamic element matching, 5, 59, 99
 - of bipolar transistors, 62
 - of capacitors, 195, 248
 - of current sources, 60, 234, 276
 - of resistors, 180
- Dynamic error correction, 8, 146, 272, 274
 - See also* dynamic error signals
- Dynamic error signals, 114, 138–139, 146–148
 - bitstream-controlled, 150
 - pseudo-random clocking, 150
 - See also* dynamic error correction
- Early effect, 36
 - forward, 37
 - reproducibility, 221, 223, 273
 - reverse, 38, 273
- EEPROM, 77
- Emitter injection efficiency, 19
- ENOB, 109
 - derivations, 277

- of first-order sigma-delta modulator, 123
- of leaky first-order sigma-delta modulator, 127
- of leaky second-order sigma-delta modulator, 137
- of second-order sigma-delta modulator, 135
- requirement, 109
- EPROM, 77
- Equivalent noise resistance, 172, 190
- Error budget, 55
- Experimental results, 235, 240, 260
- Feed-forward path, 131
- FIR filter, 138–139
- Forward-active region, 14–15, 20, 37
- Fusible links, 77
- Gain-boosting, 255
- Gummel number, 16, 21, 68
- Gummel plot, 17
- High-injection region, 17–18, 20, 34
- Incremental ADC, 120
- INL, 109
- Integration time, 166, 168
- Inter-symbol interference, 172
- Intermediate offset storage, 201
- Intrinsic carrier concentration, 13, 15, 34
 - temperature dependency, 21
- Kelvin connection, 180
- Laser trimming, 76
- Lead frame, 33
- Leakage current, 17, 41
 - errors in CT circuitry, 169
 - errors in SC circuitry, 185
 - in ESD diodes, 220
- Leakage, 125
 - in first-order modulators, 127
 - in first-order sigma-delta modulator, 278
 - in second-order modulators, 136
 - in second-order sigma-delta modulator, 281
 - in sigma-delta modulators, 125
- Limit cycles, 125, 130, 281
- Lithographic errors, 28, 30, 43
- LM75, 227
- Locking, 113
- Loop filter, 117, 119
 - averaging of errors in, 149
 - first-order, 121
 - implementation, 239
 - second-order, 130–131
- Loop gain, 67
- Mechanical stress, 5, 7, 31–32, 74, 216–217
 - causes, 33
 - due to packaging, 273
 - effect on bipolar transistors, 34
 - effect on current gain, 35
 - effect on resistors, 45
 - effect on saturation current, 34
 - temperature dependency, 75
- Military temperature range, 80
- Mismatch
 - between capacitors, 184
 - between current sources, 59, 275
 - between input impedances, 201
 - between leakage currents, 186
 - between resistors, 168, 180
 - between switches, 62, 179
 - between time constants, 172
 - between transistors, 59, 175
 - charge-injection, 197, 202
 - in cascaded sigma-delta modulators, 130
- Motivation, 1–2
- Multi-bit quantizer, 118
- Nested-chopper technique, 5, 200, 232
 - residual offset, 201
- Noise shaping, 117
 - effect of filter coefficients, 135
 - first-order, 121
 - second-order, 134
- Noise transfer function, 119
- Noise
 - in sigma-delta ADC, 162
 - of bipolar front-end, 163
 - of continuous-time implementations, 172
 - of switched-capacitor implementations, 187
- Non-ideality factor, 14, 39
- Non-overlapping clock, 239, 255
- Normal-mode rejection, 113, 138, 146
- Offset cancellation, 175
 - advanced techniques, 196
 - trimming, 175
 - See also* autozeroing, chopping
- Offset
 - causes, 175
 - due to charge injection, 196
 - in bias circuit, 66, 251
 - in CMOS technology, 5, 175
 - in continuous-time readout circuitry, 167
 - in switched-capacitor readout circuitry, 184
- Operating space, 216
- OTP, 77
- Output impedance, 62–63, 166, 168, 170
- Oversampling, 116
- Packaging shift, 33, 55, 217
- Packaging, 33, 74
 - ceramic, 33–34, 242, 262
 - metal can, 33–34
 - plastic, 5, 7, 33–34, 74, 217, 235
- Parasitic substrate pnp, 25
- Platinum resistor, 2, 217, 242, 262–263
- Polysilicon resistors, 43
- Power consumption, 160

- of continuous-time implementations, 174
 - of switched-capacitor implementations, 190
- Power-supply rejection, 64, 114, 198, 257, 266
- Production costs, 2, 7, 68, 185, 213, 217
- Production spread, 6, 215, 217
- PTAT errors, 56, 68, 75
 - due to bias-resistor spread, 43
 - due to current-gain spread, 31
 - due to saturation-current spread, 31
- Quantization error, 109
- Quantization noise, 115
- Quantization, 115
- Quantizer, 115–117
- Random errors, 55
- Ratiometric curvature correction, 89, 228, 231, 248, 262
 - higher-order, 90
- Ratiometric temperature measurement, 3, 5, 39, 51, 89
- Recombination factor, 20
- Recombination, 13, 17, 20, 37
- Remote diode, 6, 272–273
- Resistors, 40
 - bias, 63
 - correlation with saturation current, 68
 - effect of mechanical stress, 45
 - matching, 168, 244
 - metal, 41
 - non-linearity, 168
 - polysilicon, 40, 45, 77, 169
 - processing spread, 43
 - shallow diffusion, 40, 45
 - spread of temperature coefficient, 273
 - temperature coefficient spread, 44
 - temperature dependency, 42, 81–82
 - trimming, 71, 76, 236
 - well, 41
- Sampling, 115
- Saturation current, 13, 16, 52, 287
 - correlation with well resistance, 41
 - effect of mechanical stress, 34
 - processing spread, 29
 - temperature dependency, 20, 81
- Self-heating, 17, 160–161, 174, 190–191, 220, 274
- Series resistance, 26, 36
 - errors due to, 96
 - instantaneous compensation, 96
 - reducing error due to, 37
 - sequential compensation, 98, 259
- Series resistances, 17
- Settling errors, 186, 255, 283
- Sigma-delta ADC, 8, 110, 115
 - bitstream-controlled timing of dynamic error signals, 150
 - incremental operation, 120
 - initialization, 128, 240
 - normal-mode rejection, 146
 - operating principles, 115
 - per-cycle noise analysis, 162
 - See also decimation filter
- Sigma-delta modulator
 - cascading, 129–130
 - digital, 72, 250
 - feed-forward path, 131
 - first-order, 112, 114, 121, 166, 229
 - linear model, 118
 - multi-bit feedback, 129
 - resolution limits, 277
 - second-order, 129, 239, 252
 - stability, 131–132
- Signal transfer function, 119
- Sinc filter, 123
- Sinc² filter, 123
- Single-shot operation, 120
- Smart sensors, 1–2, 5, 214, 274
- Startup circuits, 64
- Switched-capacitor circuitry, 182
- Switched-capacitor integrator, 159
- Switched-capacitor loop filter, 117
- Switching transients, 166, 171, 180
- Synchronous modulators, 112
 - comparison to asynchronous, 113
- Systematic error
 - due to curvature, 79
- Systematic errors, 55
 - due to stress, 76
- Thermal capacitance, 162
- Thermal expansion, 33, 74
- Thermal management, 2, 6, 272, 274
- Thermal resistance, 161
- Thermal settling, 219
- Thermal voltage, 3, 11, 69
- Thermistors, 2, 217
- Thermopiles, 2
- Three-signal technique, 204
- Tones, 120, 242
- Traceability, 214
- Trimming, 6, 43, 69
 - after packaging, 74
 - bitstream-controlled, 257
 - current-domain, 71
 - definition, 214
 - digital, 73, 236
 - modulated, 72, 244, 249
 - non-volatile memory, 76
 - parameters, 69
 - voltage-domain, 70, 238
- Voltage monitoring, 6, 272
- Voltage reference calibration, 221–222, 244, 258–259, 264
- Voltage references, 79
- Voltage-to-charge conversion, 159, 239

Voltage-to-current converter, 166
 chopped, 176

Window function, 139

Zener diode, 77, 80
Zener zapping, 77