Low-Power High-Accuracy Smart Temperature Sensors in CMOS Technology

André Luiz Aita

Low-Power High-Accuracy Smart Temperature Sensors in CMOS Technology

Low-Power High-Accuracy Smart Temperature Sensors in CMOS Technology

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 1 maart 2011 om 15:00 uur

door

André Luiz AITA

Electrical Engineer and Master of Computer Science
Federal University of Rio Grande do Sul, Brazil

geboren te Santa Maria, RS, Brazilië

Dit proefschrift is goedgekeurd door de promotor:

Prof.dr.ir. G.C.M. Meijer

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof.dr.ir. G.C.M. Meijer	Technische Universiteit Delft, promotor
Prof.dr. S. Bampi	Federal University of Rio Grande do Sul
Prof.ir. A.J.M. van Tuijl	Universiteit Twente
Prof.dr. J.R. Long	Technische Universiteit Delft
Prof.dr.ir. A.H.M. van Roermund	Technische Universiteit Eindhoven
Prof.dr.ir. J.H. Huijsing	Technische Universiteit Delft
Dr.ir. M.A.P. Pertijs	Technische Universiteit Delft
Prof.dr. P.J. French	Technische Universiteit Delft, reservelid

This work was supported by Universidade Federal de Santa Maria, Brazil and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, Brazil, grant BEX 1877/03-8.

Printed by PrintPartners Ipskamp, Enschede

ISBN: 978-90-813316-9-2

Keywords: integrated smart temperature sensors, CMOS-compatible substrate PNP transistors, $\Delta\Sigma$ A/D converters

To my wife Letícia.
To my parents, Vinício (in memoriam) and Aldonir.

There is a legend about a bird which sings just once in its life, more sweetly than any other creature on the face of the earth. From the moment it leaves the nest it searches for a thorn tree, and does not rest until it has found one. Then, singing among the savage branches, it impales itself upon the longest, sharpest spine. And, dying, it rises above its own agony to outcarol the lark and the nightingale. One superlative song, existence the price. But the whole world stills to listen, and God in His heaven smiles. For the best is only bought at the cost of great pain... or so says the legend.

The Thorn Birds
Colleen McCullough

Contents

CHAPTER 1

Introduction

This thesis describes the theory, design and implementation of low-power high-accuracy CMOS smart temperature sensors. While there are several types of temperature sensors, the substrate PNP transistor, which is CMOS-compatible, was elected as the temperature sensor element. $\Delta\Sigma$ A/D converters on the other hand were the elected architecture for data (temperature) conversion.

These choices, discussed in detail throughout this work, took into account the main goal of this work: the development of a low-power temperature sensor, while having a state-of-the-art accuracy of $\pm 0.1°C$ (after 1-point trimming) over a wide temperature range, i.e. the military range, from $-55°C$ to $125°C$. Because of the well-known accuracy and power consumption trade-off, advanced circuits techniques were needed to circumvent the limits set by this trade-off, allowing the establishment of a very high performance temperature sensor under a very low power consumption, below $65\mu W$ at a supply voltage of 2.5V. In addition, another goal was to design a sensor with an enhanced non-trimmed accuracy (i.e. better than $\pm 0.5°C$) to be able to reduce costs by omitting the need for individual trimming while providing sufficient accuracy for many applications.

Since the design of temperature sensors is a very mature topic, a lot of effort is required to accomplish a small performance improvement. Furthermore, relatively low frequencies, up to some hundreds of kHz, are typically used in temperature sensors. This may explain why, nowadays, such theme does not seem to arouse interest of many students (when compared to other more fashionable areas)[1]. However, the design of low-power high-accuracy smart temperature sensors poses great challenges and requires indeed a deep

[1]Someone may doubt, but this statement is the author's belief.

knowledge of many instrumentation techniques (that in fact can be applied in many other areas). Temperature measurement is thus not only about measuring temperature but applying and developing instrumentation techniques and, why not, turning definitely electronics into art.

1.1 Smart (temperature) sensors

In electronics systems, a sensor can be defined as a device or circuit capable of perceiving some physical quantity, e.g. temperature, movement (speed and acceleration), position, humidity, etc, often translating this physical signal (from different energy domains) to the electrical domain [1]. Because of the analog nature of the physical quantity, the sensor is commonly combined with an analog interface (referred in this work as the sensor front-end). If this sensor is able to provide this result in a digital format, readily readable by a microprocessor or measurement instrument, by means of a data (A/D) converter or some application specific sensor interface[2], it is then classified as a smart sensor. This is schematically shown in Fig. 1.1.

While the sensor integration is normally difficult with a stand-alone sensor, even if an integrated one, it is straightforward with a smart sensor. This easy connection sensor-processing unit has allowed the implementation of low-cost but fully functional systems (with actuators too). In other words, with smart temperature sensors it is simple to collect and combine data, and based on them to make decisions, providing some type of output.

However, because of technology issues, often it is not simple to make a sensor smart, e.g. the fabrication of a smart temperature sensor for temperatures higher than 300°C requires special IC technology that can stand to such temperatures [2]. However, for an intermediate range like −70°C to 130°C, this is perfectly possible with standard technologies. Problems related to the fact that sensing elements and electronics are put together cannot be forgotten. Among these problems, the interference of the digital part to the analog one is the most common, but other related to stability requirements, size and packaging also exists.

1.2 Temperature sensors applications

Temperature and smart temperature sensors are widely applied in many areas. Some of these applications/areas are reported below:

[2]The interface can also be a general, low-cost and high-performance interface between sensing elements, named UTI (Universal Transducer Interface) [2].

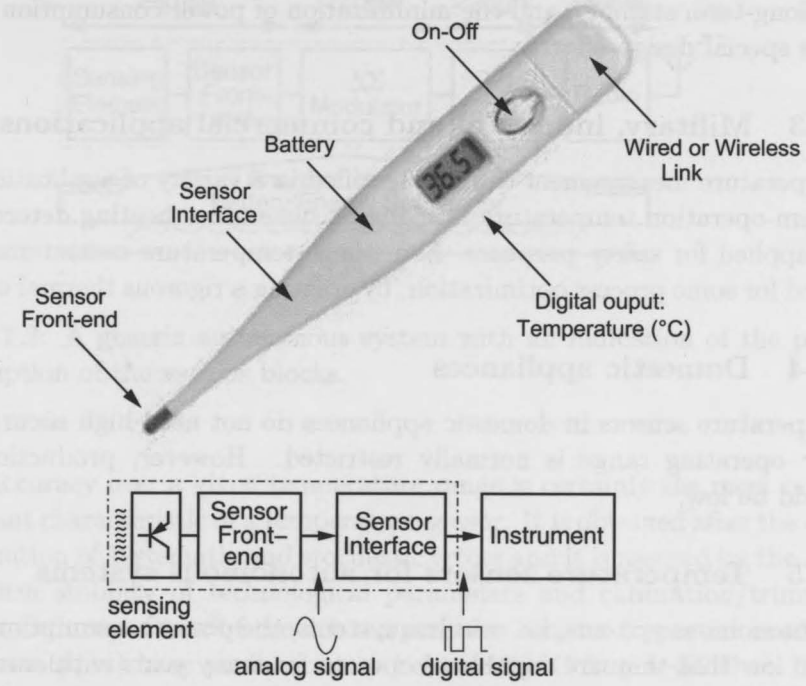


Figure 1.1: Schematic of a smart sensor. The sensor interface can be either a specific application sensor interface, or an UTI or an A/D converter. The instrument can be either a measurement instrument, a micro-controller or a digital processing unit.

1.2.1 Sensing applications

Obviously the primary application of a temperature sensor is temperature measurement, either absolute or relative, or both, simultaneously. But temperature sensors are used also for measurement of other quantities, a process called temperature translation [3]. In addition, because many sensors show some degree of temperature sensitivity (an effect called cross-sensitivity), and because they cannot be shielded against temperature variations, temperature sensors can be used to measure their temperature, so to make some type of compensation possible [4].

1.2.2 Biomedical and clinical (medical) applications

Temperature measurement is also important for clinical evaluation. The restricted temperature range certainly eases the design, but the reliability

and long-term stability, and the minimization of power consumption require often special design efforts.

1.2.3 Military, industrial and commercial applications

Temperature measurement can be identified in a variety of applications, e.g. system-operation temperature tracking or system-overheating detection, often applied for safety purposes. Sometimes temperature measurements are aimed for some process optimization, by creating a rigorous thermal control[3].

1.2.4 Domestic appliances

Temperature sensors in domestic appliances do not need high accuracy and their operating range is normally restricted. However, production costs should be low.

1.2.5 Temperature sensors for autonomous systems

In autonomous systems, i.e. wireless systems, the power consumption should be so low that they are capable to operate for many years with energy supplied from a battery or with energy harvested from the environment. Such systems are becoming more and more important in industry, either in a stand-alone configuration or in sensor networks [5]. Low-power smart temperature sensors are rather suited as one of the fundamental building blocks of such systems[4] and can be operated together with energy scavengers and transmission-efficient radios[5].

Figure 1.2 shows a typical autonomous system, where a smart sensor (delimited by a dashed square) can be identified. Examples of such autonomous systems are the Tire Pressure Monitoring Systems (TPMS) [6] as well sensor nodes used to implement sensor networks [5].

1.3 General characteristics of a sensor

The main characteristics of a temperature sensor[6], which are often used as design specifications, are summarized below [3][7]:

- *Accuracy*

[3]Such applications sometimes are referred as thermal management applications.

[4]Such systems can have an entire processing unit inside, but in most of the cases, some type of sensor, e.g. temperature sensor, is present.

[5]The study of radios and energy scavengers is out of scope of this work.

[6]These characteristics are also valid for many other sensors.

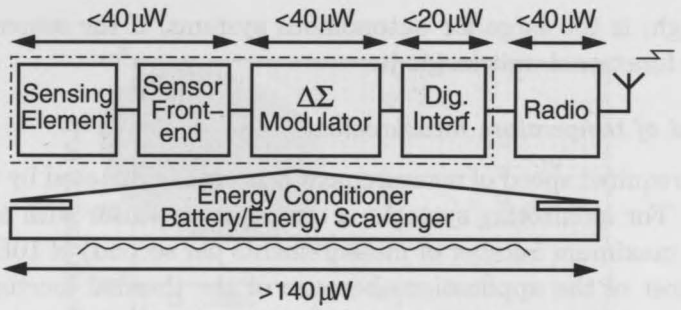


Figure 1.2: A generic autonomous system with an indication of the power consumption of the various blocks.

Accuracy over a given temperature range is certainly the most important characteristic of a temperature sensor. It is obtained after the elimination of systematic and stochastic errors and it is assured by the long-term stability of technological parameters and calibration/trimming procedures. State-of-the-art temperature sensors (reported academic prototypes) have an inaccuracy below $\pm 0.1°C$ from $-55°C$ to $125°C$ after trimming at room-temperature [8] and a batch calibrated (non-trimmed) inaccuracy below $\pm 0.25°C$ from $-70°C$ to $130°C$ [9].

- *Temperature range*

  The temperature range is usually referred as the operating range within which the accuracy is guaranteed. Some well-known ranges are the commercial, industrial and military ranges, ranging from $0°C$ to $85°C$, from $-40°C$ to $100°C$ and from $-55°C$ to $125°C$, respectively. The clinical (medical) range is also frequently used and ranges from $35°C$ to $42°C$.

- *Power consumption*

  In the past, the power consumption was not an important characteristic, probably being much more related to the sensor self-heating than to the power per conversion itself. The main concern was to limit the self-heating, so to be able to measure temperature, accurately. Nowadays, the power consumption may have the same status of accuracy, or, as in the case of some portable battery supplied applications or autonomous systems, more importance at all. Typical power consumption, as given in Fig. 1.2, of newer temperature sensors are below $140\mu W$, for accuracy levels as mentioned in the former item accuracy. This value,

though, is too large for autonomous systems, if the sensor is not the only functional unit inside it.

- *Speed of temperature measurement*

  The required speed of measurement is normally dictated by the application. For monitoring systems, a temperature sensor with a bandwidth (i.e. maximum number of measurements per second) of 10Hz is enough in most of the applications because of the thermal inertia. However, in control systems, for stability reasons, a much higher bandwidth is often required.

- *Adaptability*

  Adaptability concerns the sensor's ability to adapt itself to the application requirements and constraints while in operation, e.g. reducing the power consumption per conversion (in case of lack of battery energy) by modifying (reducing) the data measurement rate and or measurement speed.

- *Cost*

  The development of temperature sensors with low-cost CMOS technologies, which allows signal conditioning and digital signal processing on the same chip, and availability of a high accuracy without trimming (or at most after a single trimming step) are often two requirements for a low-cost temperature sensor.

- *Long-term stability and Electro-Magnetic Susceptibility (EMS)*

  While the long-term stability characteristic predicts the sensor's accuracy behavior in the long run, Electro-Magnetic Susceptibility or interference robustness refers to the measurement reliability in the presence of any internal[7] or external interference.

1.4 State-of-the-art of smart temperature sensors

Figure 1.3 presents a temperature sensor performance survey with collected data from 1996 to 2010 [10], where the energy per conversion versus relative error is discussed. A Figure-of-Merit (FOM) given by the energy per conversion (nJ) versus relative error^2 (%)^2, as defined and presented in [10], is

[7] In mixed-signal sensor design, the digital circuitry is source of many spurious signal that may cause a potential interference over the sensitive analog parts of the sensor.

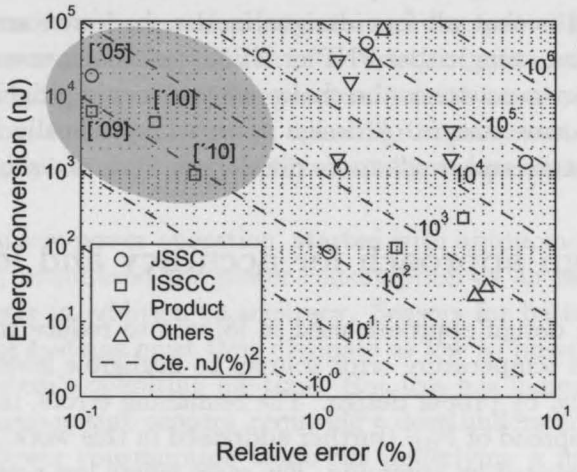


Figure 1.3: Temperature sensor performance (energy/conversion versus relative error) survey from 1996 to half of 2010. Obtained from [10]. Dashed lines indicate constant FOM (energy per conversion (nJ) versus relative error^2 $(\%)^2$). The shadowed area shows the most accurate sensors up to 2010 (the year of publication of these sensors is indicated in square brackets).

used for comparison, what in fact agrees with the FOM presented in [11], as discussed in the Section 3.8.4 of Chapter 3.

Figure 1.3 shows that since 2005, the smart temperature sensor described in [8] (named ['05] in the picture) represents the state-of-the art in respect to accuracy, having it an inaccuracy below $\pm 0.1°C$ from $-55°C$ to $125°C$ (i.e. a relative error[8] of 0.11%). The work presented in [9] (named ['09] in the picture) and also in this thesis has a state-of-the-art accuracy too, since it has the same relative error of [8]. However, it dissipates less, thus meaning a higher performance, i.e. smaller FOM (3 times) , as shown in the Figure 1.3.

Figure 1.3 shows that the energy per conversion can be reduced by allowing an increased error in the sensor output. The challenge is, for instance, to keep a certain accuracy level but under a lower power consumption and thus FOM. For instance, the most accurate sensors are in the top-left corner of the picture, (the shadowed area of Figure 1.3), where the energy per conversion is typically high. Very recent works [12][13] (both named ['10] in the picture) report a lower energy per conversion, but at expense of a higher relative error in regard to the state-of-the art [8][9].

[8]The relative error is obtained dividing the maximum error in the temperature range by the temperature range, e.g. $\frac{0.2}{180}$ =0.11%.

Another interesting and important observation that can be obtained from the survey [10] is that all four designs in the shadowed area are based on bandgap circuits, using either PNP or NPN transistors, revealing that up to now the bipolar transistor is the choice when designing highly-accurate low-power temperature sensors. A design approach specifically for this purpose is not fully mature and needs to be developed. This is discussed as follows.

1.5 Design approach for accuracy and low-power

The successful design approach used in [8] was to reduce the effect of each of the different temperature error sources to negligible levels, i.e. $\frac{1}{10}$ of the overall accuracy, by proper design. The remaining errors, in particular that related to the spread of V_{BE} (further addressed in this work), not avoided by design, were reduced by trimming. For cases where low-power consumption is required too (i.e. the sensor has some specified power constraint), this approach may result in a sensor whose implementation contains some degree of over-design (e.g. bias currents in excess given the accuracy specification).

In this sense, a modified design approach to work with both accuracy and power consumption requirements was proposed and used thoroughly this work. More specifically, in many parts of the system, relations between accuracy and power consumption were derived, and given accuracy requirements, power consumption values were found. In practice, this approach meant keeping most of the error sources bounded to a smaller (but not negligible) level than the required final accuracy (with the associated power saving), assuming that the inaccuracy would still be low, in order to be improved by trimming.

1.6 Motivation and objectives

Temperature sensors, as discussed, have many applications. In typical ones, where accuracies of about $\pm 1°C$ are needed, the design of integrated sensors is relatively easy. This is because of the current state of electronics today. However, if higher accuracy is required, the design of temperature sensors can be very challenging, requiring a number of circuit techniques to circumvent the electronics limitation and their non-idealities.

The growing trend in the field of thermal management, mainly started with the microprocessors market for portable computers, but nowadays not restricted to it, increased the demand for integrated temperature sensors and helped for their spread. Yet, pushed by the availability of highly-accurate discrete temperature sensors, like the Pt100 resistors, integrated temperature

sensors needed to be accurate too in order to find their market. Nowadays, temperature sensors appear as fundamental circuits, providing means for a safe temperature-controlled operation of a variety of systems.

Smart (with a digital output) temperature sensors came next. Since most of the applications are often based on some digital processor, it is important to provide a highly-accurate and readily-available temperature in the digital format.

The trend in low-power operation, started with autonomous systems, on the other hand, introduced the power consumption as an important specification parameter in addition to accuracy. Sensors for battery-supplied or even battery-less systems must thus dissipate as low as possible, in order to preserve the system's operating life-time. But this has increased the design complexity of temperature sensors, requiring a deep understanding of important accuracy-power consumption trade-offs underlying a more systematic design.

1.7 Organization of the thesis

This thesis is organized in six main chapters as follows.

Chapter 2 - Temperature Sensing Fundamentals presents the design background on temperature sensors. The substrate PNP transistor characteristics and behavior are discussed in detail since this transistor is the fundamental building block of the sensor.

Chapter 3 - Front-end presents the design of the analog (sensor) front-end, showing how to generate accurately important voltages like the PTAT voltage ΔV_{BE} and the CTAT voltage V_{BE}. PNP transistors bias currents and current ratios are deeply analyzed in face of accuracy requirements and power consumption constraints.

In Chapter 4 - $\Delta\Sigma$ A/D Converters, the A/D converter of the smart temperature sensor, i.e. the sensor digital interface, is analyzed and designed. This converter processes the front-end analog outputs, i.e. ΔV_{BE} and V_{BE} signals, generating a digital signal, whose value is a digital representation of the temperature. Important accuracy-power consumption tradeoffs are addressed focusing on a low-power though still accurate A/D converter.

In Chapter 5 - Realizations, measurement results of five temperature sensors are presented and analyzed. The last sensor, which is not coincidentally the sensor with best performance, having a batch-calibrated inaccuracy below $\pm 0.25°C$ ($\pm 3\sigma$) from $-70°$ C to 130°C, with a total supply current of $25\mu A$, is discussed in more details.

The thesis ends with Chapter 6 - Conclusions, where the main findings

and original contributions of this work are given and once more commented.

The Appendix brings some additional text about selected subjects of this work.

CHAPTER 2

Temperature Sensing Fundamentals

2.1 Introduction

In this chapter most of the fundamental background to on-chip temperature measurement is provided. Frequently used application-related requirements are initially discussed. After, some of the most well-known temperature sensors are described and compared. Because of the particular requirements of this work, the bipolar transistor is chosen as the temperature sensor. Therefore, it is described in more detail, so as to provide enough knowledge about its properties and operating characteristics. This include its behavior as a function of temperature, and the main causes of fluctuation (variation due to spread) in its characteristics, which are particularly important when quantifying sources of measurement errors, topic of the next chapter. Finally, substrate PNP transistors (which are available as CMOS compatible transistors) and a typical band-gap circuit are presented.

2.2 Application-related requirements

The design of a sensor, like the design of many other systems, is based on a set of specifications the application requires, often named application-related requirements. In the case of a (temperature) sensor, the most important specification is often the accuracy the sensor has over a given operating (temperature) range. Power consumption is an important specification too and may become the most important one if the system is autonomous or battery supplied, when there is a severe power (and/or energy) constraint. Other

requirements include, bandwidth, type of interface, among others.

Accuracy and resolution

Accuracy (or altenatively inaccuracy) sets the maximum error in a measuremnt over a given temperature range. In this particular work, an accuracy level of $\pm 0.1°C$ over the entire temperature range is targeted. Resolution, since it refers to the ability to recognize a small change in the input, e.g. temperature variation, is set to some extent smaller than the accuracy (refer to Section 4.1.1 of Chapter 4).

Temperature range

The temperature range is referred as the operating range within which the accuracy is guaranteed. In this work, most of the design efforts were done thinking of the military range, i.e. from $-55°C$ to $125°C$. In some cases a wider temperature range was also adopted, ranging from $-70°C$ to $130°C$, as well some measurements done up to $150°C$.

Power consumption

The power consumption of a sensor is an important specification, specially when the sensor is thought for some portable battery supplied application or autonomous systems.

In this work, the total sensor power consumption budget was set to below $75\mu W$. This value was based on typical power consumption of current low-power temperature sensors. While certain autonomous applications require a much smaller power consumption, this value, given the specified accuracy accuracy level of $\pm 0.1°C$ over $180°C$, presents relevant performance increase with respect to the prior art [8].

Speed of measurement and bandwidth

A bandwidth of 10Hz (i.e. 10 measurements per second) was set. This bandwidth is enough for most of the applications (whether some applications might require a higher bandwidth, other specifications are often relaxed, e.g. power consumption or temperature range). In particular cases, it is important to have a programmable conversion rate, e.g. given an energy constraint, a conversion rate smaller than the full bandwidth might be appropriate.

Sensor interface

Smart sensors are required to provide the information they generate in a digital format. An A/D converter is thus often needed. In the case of temperature sensors, where the signal has a very small bandwidth (tens of Hz), oversampled A/D converters, e.g. $\Delta\Sigma$s, are preferred. Such data converters have the ability to self-adapt to the application needs, not to mention the relaxed specifications for the analog components [14].

2.2.1 Other characteristics

Immunity to interference

Immunity to interference or interference robustness is often required from a sensor. $\Delta\Sigma$ architectures have one important advantage against their Nyquist-rate counterparts: immunity against digital interference[1]. Because of the oversampling, the effect of any interference (within a $\Delta\Sigma$ cycle) is attenuated in regard to the full conversion.

Self-heating

Self-heating refers to the condition when the sensor has its internal temperature increased by its own operation. Since this condition cannot be avoided, self-heating has to be small so to limit temperature measurement errors. In this sense, the most effective way of handling this condition is then by limiting the sensor power consumption to below, for instance, a few hundreds of μW to avoid introducing errors greater than 0.05°C in typical IC ceramic packages [15]. Operating the sensor periodically (i.e. duty-cycled operation as follows) might be another solution to reduce self-heating. However, this will affect the sensor availability.

Duty-cycled operation

Because most of the sensor signals have a very small bandwidth, the sensor can operate duty-cycled [16]. After a measurement, the system is switched-off. This non-continuous or duty-cycled operation is important mainly in autonomous sensors when the available energy is limited. The main consequences of a duty-cycled operation are:

- energy saving that prevents battery discharge and allows longer operation time;

[1] In fact, a property of integrating architectures

- reduction of self-heating effects (as mentioned) and;
- improved compatibility in multiplexed systems.

Another consequence of a duty-cycled operation concerns the transient time between the switched-off mode (or sleeping mode) and the operating mode, which is well known as wake-up time. While in this particular work, this time (tens of μs) can be neglected given the low frequencies involved in temperature measurements, it may be of great concern in other cases due to the amount of energy consumed during the wake-up time.

2.3 Temperature sensors

In theory, any element with temperature-dependent characteristics can be used to sense temperature. However, such characteristics are not always well-behaved, e.g. linear with temperature and only dependent on temperature, exhibiting no cross-sensitivity. This severely restricts the applicability of many elements as temperature sensors. Nonetheless, a wide range of temperature sensors is still available. The most important ones are [2]:

- transistors (mostly bipolar transistors);
- thermocouples;
- platinum (Pt) resistors;
- thermistors .

A summary with the main characteristics of each one is presented in Table 2.1. Based on these characteristics, given an application and requirements, one can define the most suitable temperature sensor.

Regardless of the type, these sensors are available as individual components, i.e. in discrete form, or built together with integrated circuits, i.e. in integrated form. All types listed in Table 2.1 are usually available in the discrete form, while some only in the integrated form. The reason is that, in general, standard CMOS integration technologies are not fully compatible with materials/procedures required to build certain types of temperature sensors, requiring additional steps, which often increase the production costs.

Discrete platinum resistors (e.g. Pt100) are very stable and because of that were listed as the interpolating temperature standard from 13.9K to 962°C according to the International Temperature Scale [17][2]. Because this

[2]More than one Pt resistor is required to provide reasonable accuracy over this wide temperature range.

Table 2.1: Characteristics of the main temperature sensing elements. Obtained from [2].

Characteristics	Transistors	Thermo-couples	Platinum resistors	Thermistors
Temperature range (°C)	medium [-80, 180]	very large [-270, 3500]	large [-260, 1000]	medium [-80, 180]
Absolute accuracy	medium	no	high	high
Accuracy for small temperature variations	medium	high	medium	medium
Integrability in standard CMOS	yes	yes	no	no
Sensitivity	high (2mV/K)	low (1mV/K)	low (0.4%/K)	high (5%/K)
Linearity	good	good	good	strongly non-linear

range includes the particular range of interest of this work, i.e. −55°C to 125°C, namely the military range, calibrated Pt100 resistors were used to calibrate the integrated sensors described in this thesis.

Thermocouples offer better performance than Pt resistors at higher temperatures, but because of their low cost and high reliability, they also find use in other temperature ranges. The main issue of a thermocouple is that it generates a voltage that is proportional to the difference between the hot and cold junction of the element, providing only a relative temperature measurement. A second absolute temperature sensor is usually required to measure the temperature of the cold junction.

Thermistors are very low cost and very compact in size. They can be also very sensitive, exhibiting thus excellent resolution, but over temperature they present a strong non-linearity behavior, which imposes extra complexity to process the signal.

Transistors are as sensitive as thermistors and have the advantage that the base-emitter voltage is almost linearly related to the temperature.

2.3.1 Integrated temperature sensors

Integrated sensors are increasingly becoming preferred since merging the sensor with the processing circuit in the so called smart sensor allows compact

and low cost sensors, not to mention an output in a digital format, ready for use.

Thermocouples and transistors are the two best examples of temperature sensors that can be integrated. Although it is possible to make Pt resistors on chips, this is not a standard procedure up to now in CMOS technology, meaning increased fabrication costs. Not to mention that thin-film resistors are not as good as their discrete counterparts [2].

Thermocouples are intrinsically offset-free and certainly the best option when the temperature is higher than transistors can stand (250°C). But the fact that they only measure relative temperature is a serious disadvantage.

Temperature sensors based on thermal diffusivity, named ETFs (Electrothermal Filter) [18][19] can also be successfully integrated. But they have a large power dissipation, due to an internal heating source, and this severely impairs the design of low power sensors. Moreover, there is need for a reference frequency, which is in general not available.

Bipolar transistors (PNPs and NPNs) are thus the most natural option for integration and the most suited temperature sensor if absolute temperature measurement is required and low-power consumption is needed. A reference voltage V_{REF} is required in this case, but that can be successfully derived from the band-gap voltage $V_{bg} \approx 1.2V$ of the Silicon at 0K. Adding to this, they are also stable and sensitive, with a reasonably linear behavior.

Although MOS transistors can be used to implement temperature sensors as well as bipolar transistors [20], mutual compensation of mobility and threshold voltage temperature effects is required [21], which adds to design complexity. Without such compensations, the MOS transistor is typically much less accurate.

2.3.2 Platinum (Pt) resistors

Platinum resistors (more specifically Pt100 resistors) were used throughout this work as temperature references. They are briefly mentioned in this section for the sake of completeness. A detailed discussion of the properties of thermocouples and thermistors is beyond the scope of this thesis.

The relation between resistance and temperature in Pt resistors is generally expressed as (DIN-IEC 751 Standard) [2]

$$R(T) = R_{Pt}(T_0)(1 + a_1T + a_2T^2 + a_3T^3 + a_4T^4)$$

where $R_{Pt}(T_0)$ is the nominal resistance at temperature $T = T_0 = 0°C$ and a_i the coefficients obtained after the calibration[3].

[3]The calibration is done at authorized calibration institutes. Different coefficients might be required depending on the temperature range.

Alternatively, the temperature can be expressed as

$$T = (b_0 + b_1 R + b_2 R^2 + b_3 R^3 + b_4 R^4 + ... + b_i R^i)$$

where R is the measured resistance and b_i are the inverse coefficients.

2.4 Bipolar transistors

The bipolar transistor is the most widely used building block of integrated smart temperature sensors. This is because it provides information, not about the relative, but about the absolute temperature. However, generating, reading and processing this information is usually not simple and often care must be taken. That is why it is important to know the main characteristics of such transistors and to understand their operation. With such knowledge, a proper transistor model can be formulated and used throughout the sensor design. The characteristics of bipolar transistors will be discussed extensively in the following subsections. Luckily, much of the work was already done [22][23][24][15].

2.4.1 The ideal $I_D - V_D$ diode characteristic

Most temperature sensors make use of the bipolar transistor in a configuration that is well-known as a diode-connected (Fig. 2.1). In this case, the bipolar transistor, a three terminal device, ends up having effectively only two ones, and having an operating behavior very similar to a diode. Because of the configuration, the original PNP junctions of a bipolar transistor behave somewhat like a single PN junction, which is exactly a diode. Before briefly introducing its characteristic in this section, it is important to mention that the discussion in this work, as well in many others, is formulated for a PNP transistor rather than for a NPN, since the best performance bipolar transistors available in most CMOS processes are the vertical (or substrate) PNP transistors (refer to Subsection 2.4.9 for more details). A last remark concerns the sign convention. In order to keep the same convention of former works in this area (see for instance [22]), V_{BE} and ΔV_{BE} nomenclatures are used instead of V_{EB} and ΔV_{EB}[4].

The well-known ideal diode characteristic is given by

$$I_D = I_S \left(\exp\left(\frac{qV_D}{kT} \right) - 1 \right) \tag{2.1}$$

[4]For sake of clarity, for a PNP transistor, the emitter to base voltage V_{EB} (referenced as V_{BE}) through out this work is positive, so to set the transistor in its forward-active region, unless otherwise specified.

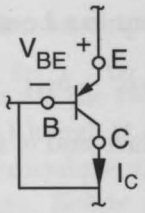


Figure 2.1: Bipolar transistor (PNP) in a diode-connected configuration.

where V_D is an external voltage applied to the device (to its p-n junction) while I_S is the saturation current. When the diode is in forward-bias (and $V_D \gg \frac{kT}{q}$), the term -1 in (2.1) can be neglected:

$$I_D \approx I_S \left(\exp\left(\frac{qV_D}{kT} \right) \right). \tag{2.2}$$

In a similar way, a voltage drop V_D that develops across the diode when a bias current I_D is applied to it is

$$V_D = \frac{kT}{q} \ln\left(\frac{I_D}{I_S} \right). \tag{2.3}$$

Remarkably, if I_D and I_S are known, by reading the voltage V_D is possible to determine the absolute temperature T. Unfortunately, I_S is strongly process (and temperature, i.e $I_S(T)$) dependent and cannot be determined.

However, if two bias currents I_{D1} and $I_{D2} = mI_1$ are applied to two diodes (assumed matched, i.e. $I_{S1} = I_{S2}$), the difference in voltage drop is

$$V_{D2} - V_{D1} = \frac{kT}{q} \ln\left(\frac{I_{D2}}{I_{S1}} \right) - \frac{kT}{q} \ln\left(\frac{I_{D1}}{I_{S2}} \right) = \frac{kT}{q} \ln(m). \tag{2.4}$$

In this case, it is not a voltage drop V_D but the difference of two voltage drops that is proportional to absolute temperature. The advantage is that the difference is not dependent on the saturation current or any other process-related parameter. This remarkable result can be used to measure temperature.

However, non-ideal effects not modeled in (2.1) make diodes not really appropriate for accurate temperature measuring [22]. Because (2.1) does not take into account the current due to generation and recombination phenomena in the depletion region of a diode, it is only approximate. Since the latter current has a different temperature dependency than that given in (2.1), the diode current cannot be expressed in a single exponential. In practice, the following approximation

$$I_D \approx I_S \left(\exp\left(\frac{qV_D}{nkT} \right) - 1 \right) \tag{2.5}$$

is used to model the diode current, initially given by (2.1), where n is the so-called non-ideality factor.

The factor n varies from 1, for large forward-bias V_D, when diffusion dominates, to 2, for low forward-bias, when recombination dominates. With this approximation, the PTAT voltage given by (2.4) ends up depending on n, which is, like I_S, process dependent. This is hardly acceptable in accurate temperature sensors[5].

2.4.2 The ideal $I_C - V_{BE}$ bipolar transistor characteristic

In a bipolar transistor, the relation between the collector current I_C and the base to the emitter voltage V_{BE} has the same form as (2.1). But because of the base terminal, the currents that are generated in the base-emitter junction of a bipolar transistor due to e.g. generation/recombination and reverse injection, as in a diode, are not part of the collector current [23], since the resulting non-ideal components of the emitter current are mainly provided via the base of the transistor [23][22]. This makes the relation $I_C - V_{BE}$ of a bipolar transistor much more ideal than the $I_D - V_D$ characteristic of a diode, i.e. n is closer to 1. That is why, despite some equivalence, a bipolar transistor has a superior performance as a temperature sensor when compared to the more simple diode.

Consider a PNP transistor, like shown in Fig. 2.2(a), in its forward-active region, i.e. with a forward-biased base-emitter junction and a reverse-biased base-collector junction. The collector current I_C can be written as [25]

$$I_C = I_S \left(\exp\left(\frac{qV_{BE}}{kT} \right) - 1 \right) \tag{2.6}$$

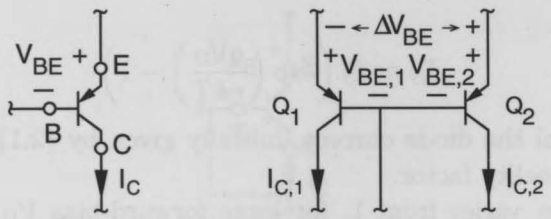
where the saturation current I_S is given by

$$I_S = \frac{qA_{BE}n_i^2\bar{D}_p}{Q_B} = \frac{kTA_{BE}n_i^2\bar{\mu}_p}{Q_B} \tag{2.7}$$

where A_{BE} is the base-emitter junction area, n_i is intrinsic carrier concentration in the base, $\bar{D}_p$ is the effective hole diffusion constant, $\bar{\mu}_p$ in the average hole mobility in the base and Q_B is the total base doping per unit area[6].

[5] Whether n might exhibit low process variation in a given technology and have its value well characterized, it is preferable to avoid any process dependency in the bias current.

[6] Q_B is well-known as the Gummel number.



(a) (b)

Figure 2.2: (a) Schematic of a PNP transistor and (b) circuit configuration to have a ΔV_{BE} voltage.

The Einstein relation $\mu = \frac{q}{kT} \cdot D_n$ was used to write I_S in terms of n_i^2 and $\bar{\mu_p}$. Again, if $V_{BE} \gg \frac{kT}{q}$, i.e. $I_C \gg I_S$, then (2.6) can be simplified to

$$I_C = I_S \left(\exp \left(\frac{qV_{BE}}{kT} \right) \right) \tag{2.8}$$

which is called the ideal $I_C - V_{BE}$ characteristic. Solving this equation for V_{BE} results in

$$V_{BE} = \frac{kT}{q} \ln \left(\frac{I_C}{I_S} \right) \tag{2.9}$$

which agrees with the diode equation given in (2.3).

The ideal logarithmic characteristic described in (2.9) shows that a PTAT voltage can be built in the same way as with diodes (and thus used to measure temperature by comparing it with a reference or temperature-independent voltage). Two equal transistors Q_1 and Q_2, as shown in Fig. 2.2(b), having collector currents $I_{C,1}$ and $I_{C,2} = mI_{C1}$, present a base-emitter voltage difference ΔV_{BE} given by

$$\Delta V_{BE} = V_{BE,2} - V_{BE,1} = \frac{kT}{q} \ln (m) . \tag{2.10}$$

This result is remarkable since ΔV_{BE} is not dependent of I_S like V_{BE} is. This makes ΔV_{BE} not dependent on process variation. As mentioned already by [26][27], the ΔV_{BE} voltage is then accurate, provided mismatch errors are eliminated.

Error due to mismatch in the saturation current I_S

If the saturation currents $I_{S,1}$ and $I_{S,2}$ of two identically designed PNP transistors are different due to a mismatch between the transistors, then the

voltage ΔV_{BE} given in (2.10) is now given (assuming for now no mismatch in the current sources):

$$\begin{aligned} \Delta V_{BE}(T) &= \frac{kT}{q} \ln \left(\frac{\frac{I_{bias,2}}{I_{S,2}}}{\frac{I_{bias,1}}{I_{S,1}}} \right) \\ &= \frac{kT}{q} \ln \left(m \frac{I_{S,1}}{I_{S,2}} \right). \end{aligned} \tag{2.11}$$

Writing $I_{S,2} = I_{S,1} \pm \Delta I_S$ results in

$$\Delta V_{BE}(T) \approx \frac{kT}{q} \ln(m) \mp \frac{kT}{q} \ln \left(\frac{\Delta I_S}{I_{S,1}} \right) \qquad (\Delta I_S \ll I_{S,1}). \tag{2.12}$$

The mismatch modifies the ideal ΔV_{BE} by the second term in (2.12). The error is proportional to the temperature (PTAT) for a temperature independent $\frac{\Delta I_S}{I_{S,1}}$.

2.4.3 The effect of I_S in the ideal $I_C - V_{BE}$ characteristic

The ΔV_{BE} voltage is accurate provided mismatch errors are eliminated. Although acceptable for relatively high bias currents, errors are no longer negligible for low bias currents. When the bias current gets small, I_S cannot be neglected, i.e. the term -1 in (2.6), and the difference ΔV_{BE}, initially given in (2.10) becomes

$$\Delta V_{BE} = V_{BE2} - V_{BE1} = \frac{kT}{q} \ln \left(\frac{m I_{C,1} + I_S}{I_{C,1} + I_S} \right) \tag{2.13}$$

in the case of two transistors with equal saturation currents, i.e. $I_{S,1} = I_{S,2} = I_S$.

This result shows unfortunately a ΔV_{BE} dependent on I_S (and now susceptible to process variations), no matter how matched the PNP transistors are nor how accurate the current ratio m is. In fact, I_S degrades the current ratio m which in turn degrades the accuracy that can be achieved from ΔV_{BE} and from the sensor. In order to prevent or limit the effect of I_S, so an accurate PTAT voltage can be obtained, the collector currents $I_{C,1}$ and $I_{C,2}$ have to be significantly larger than I_S (how large will depend on the accuracy requirements for ΔV_{BE}). Since power reduction is normally achieved by reducing the PNP bias currents, accuracy requirements for ΔV_{BE} can severely

limit power-saving attempts[7]. Requirements are much worse at high temperature, since I_S, as well-known, increases exponentially with temperature (see Section 2.4.5).

2.4.4 The effect of β in the ideal $I_C - V_{BE}$ characteristic

The collector terminal of a vertical PNP transistor (available in most CMOS processes) is formed by the p-substrate. Since the substrate is grounded (see subsection 2.4.9), there is no opportunity to bias the bipolar transistor via its collector. In this case, it is the emitter current of a bipolar transistor that is set by a current source, instead of the collector current. This explains why in a diode-connected configuration (with the base also grounded, i.e. $V_{BC} = 0$), the base current has also to be taken into account when determining the base-emitter voltage. Since the emitter current results from the sum of the base and collector currents, the collector current can be written as

$$I_C = I_E - I_B = \frac{\beta}{1+\beta} I_E \tag{2.14}$$

where β is the common-emitter forward current gain, well known and defined as the ratio $\beta = I_C/I_B$. This results in a voltage V_{BE} given by

$$V_{BE} = \frac{kT}{q} \ln\left(\frac{\beta}{1+\beta}\frac{I_E}{I_S}\right) \tag{2.15}$$

The result in (2.15) shows that a bipolar transistor biased through its emitter, even if by an ideal current source, has a V_{BE} voltage that is dependent on β, i.e. susceptible to process spread. This is particularly important when β is small ($\beta < 25$), which is the case of substrate PNPs in standard CMOS technologies [28][29]. Moreover, β has to be current-independent, so that an accurate emitter-current ratio will result in the same collector-current ratio. These results show that some technique for β compensation will be necessary for such high accuracy applications, since the β-dependency in the base-emitter voltage is a potential cause for low non-trimmed and post-trimmed accuracies. No need to mention that attention must be paid to the power consumption of the circuit used to implement the compensation.

2.4.5 Temperature Characteristic of V_{BE}

The temperature characteristic of the base-emitter voltage V_{BE} given in (2.9) is determined considering the temperature dependence of the saturation current I_S and the collector current I_C. Additionally, because $I_C = \frac{\beta}{1+\beta} \cdot I_E$,

[7] How severe this limitation is depends on the total current consumption of the sensor with regard to the total bias current of PNPs needed to generate ΔV_{BE} .

the temperature dependency of β has also to be taken into account. The analysis of I_S is carried out by analyzing its components: n_i and $\bar{\mu_p}$.

The intrinsic minority-carrier concentration n_i is given by [30]

$$n_i^2 \propto T^3 \exp\left(\frac{-qV_g}{kT}\right) \tag{2.16}$$

where V_g is the silicon band-gap approximated by

$$V_g \approx V_{g0} - \alpha T \tag{2.17}$$

where V_{g0} is the extrapolated band-gap voltage at 0K and α the linear coefficient of the approximation.

The average hole mobility in the base $\bar{\mu_p}$ is given by

$$\bar{\mu_p} \propto T^{-n}. \tag{2.18}$$

The substitution of (2.16) and (2.18) in the saturation current given in (2.7), results in[8]

$$I_S(T) = I_{S0} T^{\eta} \exp\left(\frac{-qV_{g0}}{kT}\right) \tag{2.19}$$

where I_{S0} and $\eta = 4 - n$ are constants. Equation (2.19) shows how strongly I_S depends on temperature, which before hand can lead the reader to think about the effects of such dependency in the sensor performance: while the voltage ΔV_{BE} is ideally not sensitive to I_S, V_{BE}, which is used to generate the required reference voltage V_{REF} needed for temperature measurement (as further explained in Chapter 3), is.

The substitution of (2.19) in (2.8) results in the following collector current expression:

$$I_C = I_{S0} T^{\eta} \exp\left(\frac{q}{kT}(V_{BE} - V_{g0})\right) \tag{2.20}$$

or, if solved for V_{BE},

$$V_{BE}(T) = V_{g0} + \frac{kT}{q} \ln\left(\frac{I_C}{I_{S0} T^{\eta}}\right). \tag{2.21}$$

now written explicitly in terms of temperature. i.e. $V_{BE}(T)$. Interesting to mention that because $\frac{I_C}{I_{S0}T^{\eta}} < 1$, the value of ln is negative, which gives the $V_{BE}(T)$ a negative temperature coefficient [24], despite of what can be initially misunderstood from (2.9). In this sense, the signal V_{BE} is complementary to absolute temperature (CTAT) .

[8] Q_B is assumed approximately constant with temperature [23].

In practice, the collector current is not constant over temperature but proportional to a power of T. This temperature dependence is introduced by writing I_C as

$$I_C(T) = I_{C0}T^m, \tag{2.22}$$

where I_{C0} is a constant and m the temperature exponent, e.g. a constant $I_C = I_{C0}$ is represented by $m = 0$ and a PTAT current $I_C = I_{C0}T$ represented by $m = 1$. By replacing (2.22) in (2.21) the following results

$$V_{BE}(T) = V_{g0} + \frac{kT}{q} \ln\left(\frac{I_{C0}}{I_{S0}T^{\eta-m}}\right) \tag{2.23}$$

or, if written as a function of V_{BE} at a given temperature, e.g. $V_{BE}(T_r)$ [23] (see [24] for a detailed derivation),

$$V_{BE}(T) = V_{g0}\left(1 - \frac{T}{T_r}\right) + \frac{T}{T_r}V_{BE}(T_r) - (\eta - m)\frac{kT}{q}\ln\left(\frac{T}{T_r}\right). \tag{2.24}$$

As earlier found by [23], the expression for V_{BE} as given above, when using empirical values for V_{g0} and η according to the basis of physics, is not accurate. This is due to poor approximations used in (2.17) and (2.18). Curiously though, when using V_{g0} and η as fitting parameters (derived form measured V_{BE} data), accuracy as good as 0.03K can be found.

For convenience, but also to have a better comprehension of V_{BE}, often (2.24) is rewritten as the sum of a constant term V_{BE0} (temperature independent), a linear term aT (proportional to absolute temperature) and a non-linear term $nl(T)$ (the sum of all higher-order terms) [23]:

$$V_{BE}(T) = V_{BE0} + aT + nl(T) \tag{2.25}$$

where V_{BE0} is the extrapolation of the tangent at $T = 0$K given by

$$V_{BE0} = V_{g0} - nl(0), \tag{2.26}$$

a is the slope of the tangent given by

$$a = \frac{V_{BE}(T_r) - V_{BE0}}{T_r - 0K} \tag{2.27}$$

(in such way that the linear term represents a line tangent to the $V_{BE}(T)$ curve at $T = T_r$) and $nl(T)$ the V_{BE} curvature given by

$$nl(T) = (\eta - m)\frac{K}{q}\left[T - T_r + T\ln\left(\frac{T_r}{T}\right)\right] \tag{2.28}$$

which is, at $T = 0$K, equal to

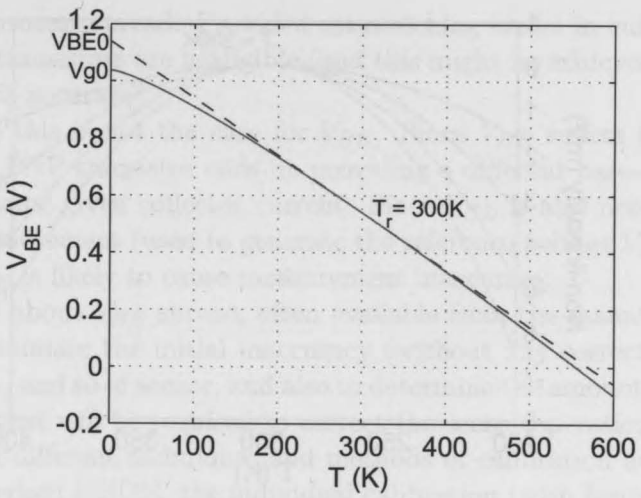


Figure 2.3: The base-emitter voltage V_{BE} as a function of temperature T. The dashed line shows a linear approximation for V_{BE} (tangent to the V_{BE}curve at $T = Tr$.

$$nl(0) = (\eta - m)\frac{K}{q}[-T_r]\,. \tag{2.29}$$

Figure 2.3 shows the base-emitter voltage $V_{BE}(T)$ as given in (2.25) versus T from 0 to 600K for $V_{BE0} = 1.15$V and $a = -2m\text{V}/°\text{C}$ ($V_{BE}(T_r = 300\text{K})$= 0.54V) while the non-linearity $nl(T)$ is plotted in Fig. 2.4 versus T from 200K to 400K for values of $\eta - m$ from 1 to 4. The curvature of V_{BE} decreases when the bias current increases faster with temperature, sort of canceling the effect of the saturation current. For instance, the non linearity obtained for $\eta - m = 3$ (case of PTAT bias current I_C) is smaller than for $\eta - m = 4$ (constant bias current), which indicates that a bias current dependent on temperature can reduce the curvature of V_{BE}. The curvature amounts to some mV in the range of 200K to 400K for all cases.

2.4.5.1 Temperature Characteristic of Current Gain β

Because the vertical PNP transistor is biased via its emitter, its collector current is a fraction $\frac{\beta}{1+\beta}$ of the applied emitter current as given in (2.14), which in turn results in a β-dependent voltage V_{BE} (see equation(2.15)). Because often β happens to be temperature dependent [25], this has to be accounted when determining the temperature characteristic of V_{BE}, initially found in (2.15).

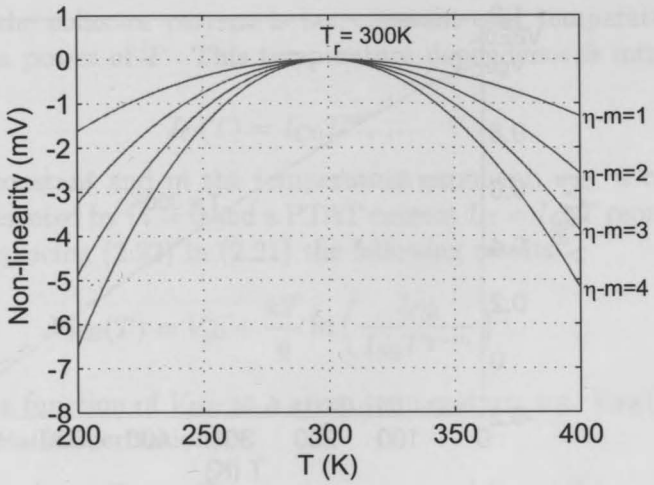


Figure 2.4: (b) The non-linearity of $V_{BE}(T)$ as a function of temperature, for values of $\eta - m$ from 1 to 4.

In practice, the following expression is used to model β [31]:

$$\beta(T) = \beta_0 \left(\frac{T}{T_r}\right)^{X_{TB}} \tag{2.30}$$

where the nominal current gain β_0 and the temperature exponent X_{TB} are found by fitting the equation to the measured data. The substitution of equation (2.30) in equation (2.15) results in a V_{BE} with an additional term

$$V_{BE} = V_{BE}(T)|_{\beta=constant} + \frac{kT}{q} \ln \left(\frac{(1+\beta_0)\left(\frac{T}{T_r}\right)^{X_{TB}}}{1+\beta_0\left(\frac{T}{T_r}\right)^{X_{TB}}} \right). \tag{2.31}$$

This additional term was studied in [22], where simulation results were presented. The adding term is only important for small values of β and large value of X_{TB}. Unfortunately, this is exactly the case for vertical PNP transistors. Nominal values for $\beta \approx 22$ and $X_{TB} \approx 4$ suggests that compensation for finite current gain is necessary when designing sensors for high accuracy. Not to mention the fact that β suffers from process spread, which causes spread to V_{BE}.

2.4.6 Process spread of V_{BE}

Probably the most important characteristic of ΔV_{BE} (the PTAT voltage as presented in (2.10) is the fact that for well chosen bias currents it does not

suffer from process spread. Provided mismatching errors in current sources and bipolar transistors are negligible (and this might be achieved by design), this voltage is accurate[9].

However, this is not the case for V_{BE}. Since V_{BE} suffers from process spread, each PNP transistor ends up providing a different base-emitter voltage for the same given collector current. Since V_{BE} is also needed for temperature measurement (used to generate the reference voltage V_{REF}), uncertainty in V_{BE} is likely to cause measurement inaccuracy.

Statistics about V_{BE} spread, often available from the manufacturer, can be used to estimate the initial inaccuracy (without any correction or trimming) of V_{BE}, and so of sensor, and also to determine the amount of trimming (e.g. in °C) that will be required to correct the sensor operation.

Although different techniques and methods of calibration and trimming have been devised [32][33], the individual calibration (with further trimming procedure) is the major cost contributor of a sensor, even if done once at room temperature. A batch calibration technique, when a few samples of the sensor are analyzed and the results used to estimate the performance of all remaining samples within a given a batch (i.e. intra-batch), can be an important replacement for individual trimming. Although it does not give the same performance, it has a much lower cost than individual trimming.

Regardless of the calibration/trimming method, it is important to analyze the most important sources of spread in V_{BE}, so as to look for adequate and efficient calibration/trimming procedures, or for ways of reducing (or avoiding) it, aiming for lower initial inaccuracies (which can help making batch calibration more efficient).

2.4.6.1 Spread of the saturation current I_S

The spread in the saturation current I_S of a PNP transistor is perhaps the most important source of V_{BE} spread. Because it is strongly process dependent, it is hard to estimate and to compensate. Usually the designer does not have any control over it, and has to work with the statistics available from the foundry.

In order to see how the spread of I_S affects V_{BE}, a deviation amount ΔI_S from the nominal value I_S is introduced in the equation of V_{BE} given by (2.9), which results in

$$V_{BE} = \frac{kT}{q} \ln\left(\frac{I_C}{I_S + \Delta I_S}\right) \tag{2.32}$$

[9]The voltage ΔV_{BE} does suffers from process spread. The spread of ΔV_{BE} , however, when compared to the spread of V_{BE} is negligible.

$$\approx V_{BE}|_{\Delta I_S=0} - \frac{kT}{q}\frac{\Delta I_S}{I_S} \qquad (\Delta I_S \ll I_S). \tag{2.33}$$

This shows the effect of the spread of I_S in V_{BE}. The original V_{BE} voltage is modified by a term that depends on the tolerance $\frac{\Delta I_S}{I_S}$. A more precise analysis thus requires the estimation of this tolerance, which is rather difficult, since I_S depends on many process parameters as shown in (2.7). Foundries often set tolerances of up to ±50%, which give worst case scenarios, fortunately, far from practical values. For instance, for AMIS0.7μm [28] , a V_{BE} spread of ±14mV is given, which corresponds to about ±7°C, when values below ±1.5°C are usually observed.

Equation (2.33) also shows that the V_{BE} spread is proportional to absolute temperature, if $\frac{\Delta I_S}{I_S}$ is assumed temperature independent. In such case, the spread of I_S causes the function $V_{BE} \times T$ to rotate around a fixed point at 0K. This suggests that trimming done just once (at a given temperature) is sufficient to correct the sensor operation. This is illustrated by Fig. 2.5, in a classical plot [22][23][15].

The assumption that $\frac{\Delta I_S}{I_S}$ is temperature independent is not necessarily true since ΔI_S has several sources. The one related to mechanical stress (also type of packaging) seems to be an example where this ratio is temperature dependent [34]. Fortunately, according to the same study, vertical PNP transistors are less sensitive to stress, and this might explain why for such transistors the temperature dependency of $\frac{\Delta I_S}{I_S}$ can be ignored.

2.4.6.2 Spread of the current gain β

Another source of spread in V_{BE} is the spread of the current gain β. Because the PNPs are biased via their emitters, V_{BE} ends up depending on β and suffering from spread in it. The spread in β is also rather difficult to estimate accurately since, like the spread in I_S, it is dependent on many strongly process-dependent parameters.

To see how the spread of β affects V_{BE}, a deviation amount $\Delta\beta$ from the nominal value β is introduced in the equation of V_{BE} given by (2.15), which results in

$$V_{BE} = \frac{kT}{q} ln\left(\frac{(\beta+\Delta\beta)}{1+(\beta+\Delta\beta)}\frac{I_E}{I_S}\right) \tag{2.34}$$

$$\approx V_{BE}|_{\Delta\beta=0} + \frac{kT}{q}\frac{\Delta\beta}{\beta}\frac{1}{1+\beta} \qquad (\Delta\beta \ll \beta) \tag{2.35}$$

This shows the effect of the spread of β in V_{BE}. The original V_{BE} voltage is modified by a term that depends not only on the tolerance $\frac{\Delta\beta}{\beta}$ but also on the factor $\frac{1}{1+\beta}$. Because of this, the spread in β will cause a non PTAT

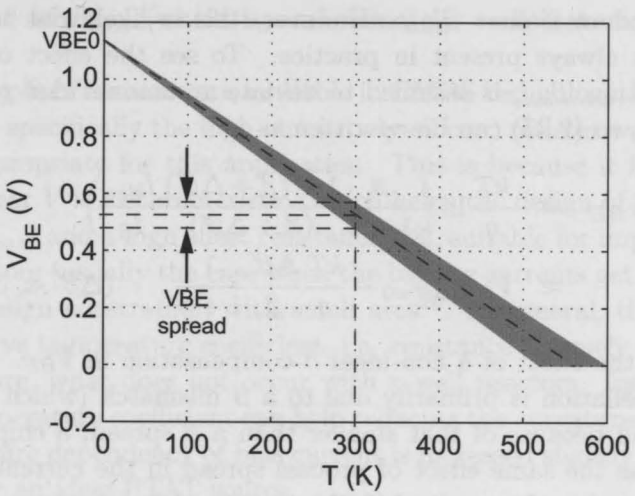


Figure 2.5: V_{BE} spread due to I_S spread (for a temperature independent $\frac{\Delta I_S}{I_S}$).

spread in V_{BE}, since β, as given in (2.30) is strongly temperature dependent. Because of the non PTAT spread, a single-point trimming method cannot correct the sensor operation successfully (simulation results in [22] shows that the remaining error in this case often presents the butterfly characteristic). Evidently, PNP transistors with larger β are less sensitive to this type of spread. Unfortunately, vertical PNPs have lack of larger current gain β (≈ 22) as already mentioned in Subsection 2.4.5.1. If, for example, $\frac{\Delta\beta}{\beta} = 20\%$ and $\beta = 22$, a V_{BE} variation of ± 0.25mV is estimated. This reinforces the need to use some technique to reduce the effect of β spread in V_{BE}.

2.4.6.3 Error due to non-ideal β-compensation

One technique, used to cancel the β-dependency of V_{BE} as given in (2.15), is designing the emitter current I_E in such way to cancel the $\frac{\beta}{1+\beta}$ factor present in the expression of V_{BE} [22]. For example, if the factor $\frac{1+\beta_{bias}}{\beta_{bias}}$ is imprinted on the emitter current I_E

$$I_E = \frac{1+\beta_{bias}}{\beta_{bias}} I_{bias} \tag{2.36}$$

used to bias the PNP transistors, the resulting V_{BE}

$$V_{BE} = \frac{kT}{q} \ln\left(\frac{\beta}{1+\beta}\frac{1+\beta_{bias}}{\beta_{bias}}\frac{I_{bias}}{I_S}\right) \tag{2.37}$$

is β-independent if $\beta = \beta_{bias}$. However, this is likely not to occur since mismatch is always present in practice. To see the effect of a non-ideal β-compensation, β_{bias} is assumed to deviate an amount $\Delta\beta'$ (due to a mismatch) of β, so (2.37) can be rewritten as

$$V_{BE} = \frac{kT}{q} \ln\left(\frac{\beta}{1+\beta}\frac{1+(\beta+\Delta\beta')}{(\beta+\Delta\beta')}\frac{I_{bias}}{I_S}\right) \tag{2.38}$$

$$\approx V_{BE}|_{\Delta\beta'=0} - \frac{kT}{q}\frac{\Delta\beta'}{\beta}\frac{1}{1+\beta} \qquad (\Delta\beta' \ll \beta). \tag{2.39}$$

This shows the effect of a non-ideal β-compensation in V_{BE}. Although the partial cancellation is primarily due to a β mismatch (which is an on-chip variation and because of that smaller than a β spread, a chip-to-chip variation), it has the same effect of process spread in the current gain (visible when compared to the result of (2.35), therefore also causing a CTAT spread of V_{BE}. Because $\Delta\beta$ is due to variation both in the nominal value β_0 as well in the temperature dependency, i.e. exponent X_{TB}, as given in (2.30), causing CTAT spread of V_{BE} regardless of the case, special attention (e.g. circuit characteristics and layout) must be taken into account.

It is worth to mention that, even under no β mismatch, the same effect can be observed, if for any reason, the β compensation scheme is not properly implemented, e.g. the bias current generated in the pre-bias circuit with the factor $\frac{1+\beta_{bias}}{\beta_{bias}}$ imprinted on it is poorly mirrored to the bias circuit (refer to details in Chapter 3). In this case, in principle, errors larger then the original ones (without β-compensation) could appear.

2.4.7 Effect of variations in the bias current

Equation (2.15) reveals a third source of spread in V_{BE} in addition to the spread due to I_S and β: the bias current variation, in this particular case, the emitter current I_E. While ΔV_{BE} is not sensitive to the bias current amplitude (and so to any variation in it), V_{BE} is. A bias current varying in amplitude affects V_{BE} in the same way I_S and β spread do.

In this work, like in [22], the PTAT bias current is set by a bias voltage, $\Delta V_{BE,bias}$ which is applied across a bias resistor R_{bias}, i.e. $I_E = \Delta V_{BE,bias}/R_{bias}$. The voltage V_{BE} as given by (2.15) is now expressed as

$$V_{BE} = \frac{kT}{q} \ln\left(\frac{\beta}{1+\beta}\frac{\frac{\Delta V_{BE,bias}}{R_{bias}}}{I_S}\right). \tag{2.40}$$

Since $\Delta V_{BE,bias}$ can be realized with relatively high accuracy (and so would be the bias current), the variation in the bias current is ultimately determined

by the spread of R_{bias} (which is, unfortunately, temperature and process dependent).

Several types of resistors are available in CMOS technology, but the poly-silicon (more specifically the high-ohmic poly-silicon) resistor was considered the most appropriate for this application. This is because it has the most relatively linear $V \times I$ characteristic (which helps the design of a PTAT bias from $\Delta V_{BE,bias}$) and a high sheet resistance [28], suitable for implementation of large resistors (usually the case when the biasing currents get small due to low-power design constraints) with small area[10]. In general, these resistors have a negative temperature coefficient, i.e. resistance inversely proportional to temperature, what does not occur with n-well resistors. Interestingly, a negative temperature coefficient can help reducing the curvature of V_{BE} since the temperature dependency of bias current is increased slightly further then that given by an ideal PTAT source.

To see how the variation of R_{bias} affects V_{BE}, a deviation amount ΔR_{bias} from the nominal value is introduced in the equation of V_{BE} given by (2.40) (the finite β is ignored for simplicity), which results in

$$V_{BE} = \frac{kT}{q} \ln\left(\frac{\Delta V_{BE,bias}}{I_S\,(R_{bias} + \Delta R_{bias})}\right) \tag{2.41}$$

$$\approx V_{BE}|_{\Delta R_{bias}=0} - \frac{kT}{q}\frac{\Delta R_{bias}}{R_{bias}}. \qquad (\Delta R_{bias} \ll R_{bias}) \tag{2.42}$$

The original V_{BE} voltage is modified by a term that depends on the tolerance $\frac{\Delta R_{bias}}{R_{bias}}$. Since the bias resistor R_{bias} is often temperature dependent (here approximated to a linear function of the temperature)

$$R_{bias}(T) \approx R_0(1 + \delta_{T1}(T - T_r)) \tag{2.43}$$

where R_0 is the nominal resistance at $T = T_r$ and δ_{T1} is the first-order temperature coefficient given by

$$\delta_{T1} = \frac{\partial R_{bias}}{\partial T}, \tag{2.44}$$

the tolerance $\frac{\Delta R_{bias}}{R_{bias}}$ can thus be expressed as

$$\frac{\Delta R_{bias}}{R_{bias}} \approx \frac{\Delta \delta_{T1} T}{1 + \delta_{T1} T} \tag{2.45}$$

[10]The sheet resistance of high-ohmic resistor is even higher than that of n-well resistor in the used technology [28].

assuming that only δ_{T1} spreads. This result shows that tolerance $\frac{\Delta R_{bias}}{R_{bias}}$ is temperature dependent. In such case, a non PTAT spread in V_{BE} is then expected due to the spread in R_{bias}. However, in general $\Delta\delta_{T1} \ll \delta_{T1}$ which indicates a small contribution to V_{BE} spread.

Additionally, because of the nominal resistor R_0 (often calculated as

$$R_0 = R_{sh}\frac{L}{W} \tag{2.46}$$

where R_{sh} is the sheet resistance and L and W the resistor length and width respectively) can also spread, a modified tolerance $\frac{\Delta R_{bias}}{R_{bias}}$ is expected. By making the resistor geometry (L, W) several times larger than the minimum given by the technology, lithographic spread can be made negligible compared to the spread in the sheet resistance. However, the spread of R_{bias} still suffers from the spread in the sheet resistance which is determined ultimately by several process parameters.

2.4.8 Other effects

There are other effects that can cause errors in V_{BE}, among them, for instance:

- base resistance: the contribution of the base resistance to V_{BE} is larger in CMOS technologies (when compared to bipolar ones). This is because of the properties of the n-well layer used to build the base of the bipolar transistor, which causes a large series base resistance R_B due to its relatively high ohmic resistance and a low forward current gain β due to its relatively large thickness. Interestingly, for a given transistor geometry, the error due to the base series resistance can be reduced by reducing the bias current which flows across the transistor. This reduces the voltage drop across R_B and consequently its effect on V_{BE}. However, the current density in the transistor is reduced too, shifting the bias operation point and transistor working region. Normally, this requires the transistor to be resized, i.e. made smaller, which in turns increases back the base resistance.

- stress: how the stress affects V_{BE} and ΔV_{BE} is fully addressed in [34]. Briefly, the vertical PNP is less sensitive to stress than the lateral one. However, it does suffer from stress, and the type of packaging often has great influence on the amount of V_{BE} spread. Trimming after packaging is therefore more appropriate, since it can correct partially the error caused by induced stress after packaging.

2.4.9 Bipolar transistors in CMOS technology

Bipolar transistors (both NPN and PNP) can be found in CMOS technologies as compatible transistors[11]. In the more common n-well CMOS processes, two types of bipolar transistors are available: lateral PNP and vertical or substrate PNP. The lateral PNP is more flexible and easy to use, since effective biasing through its collector is possible. The same does not happen with the vertical PNP, since the collector is formed by the p-substrate, which is grounded. In this case, the typical common-emitter structure which is used to generate and amplify ΔV_{BE} is not available [15], and a special amplifier configuration (OPAMP) is often required. This is shown in Fig. 2.6, where a typical circuit used to build band-gap circuits with vertical PNPs in CMOS is presented[12].

However, despite of the advantages with regard to circuit design, the lateral transistor does not perform as well as the vertical one. This is because the vertical PNP transistor, due to its physical structure (as shown in Fig. 2.7) has a more ideal $I_C - V_{BE}$ characteristic [23]. This improved behavior is fundamental when designing high performance (accuracy) temperature sensors and the drawbacks of the more complex circuitry have to be circumvented by circuit design techniques. For instance, the opamp offset V_{OS} present in the expression of the current I

$$I = \frac{V_{BE2} - V_{BE1} - V_{OS}}{R_1} = \frac{\Delta V_{BE} - V_{OS}}{R_1} \tag{2.47}$$

and in expression of the reference voltage V_{REF}

$$V_{REF} = V_{B1} + (\Delta V_{BE} - V_{OS}) \left(1 + \frac{R_2}{R_1}\right) \tag{2.48}$$

has to be removed somehow, avoiding it of jeopardizing the superior performance of the vertical PNP transistors.

2.5 Conclusions

This chapter addressed the fundamental back-ground of integrated temperature sensors design. Initially, frequently used application-related requirements of sensor were discussed. After, the most well-known temperature sensors were presented. Despite of particular advantages each one has, the

[11] The author understands that the term compatible is more appropriate than the term parasitic.

[12] This same circuit, with minor changes, might be used to generate a PTAT bias current.

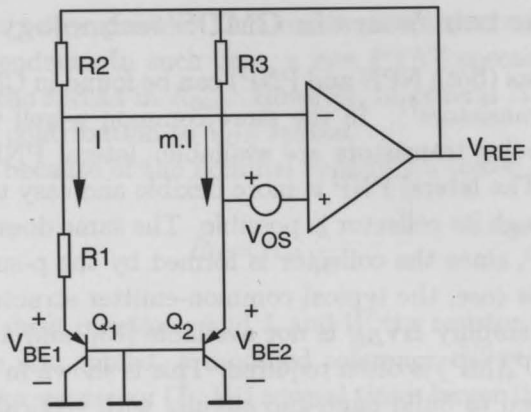


Figure 2.6: Typical band-gap circuit in CMOS technology with vertical PNP transistor.

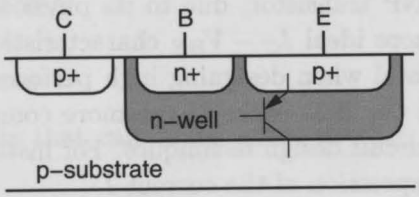


Figure 2.7: Cross section of a PNP transistor available in n-well CMOS processes.

bipolar transistor (substrate PNP) was considered the most suitable sensing element for the temperature sensor.

A detailed analysis of the bipolar transistor was thus carried out. Its ideal $I_C - V_{BE}$ was first introduced, followed by the analysis of the effect of the saturation current I_S and forward-current gain β on it. Since the PNP transistor is biased via its emitter, the ideal $I_E - V_{BE}$ characteristic was also analyzed, with further explanation of the β-compensation technique.

The temperature characteristic of V_{BE} was then presented, and its non-linearity over temperature studied. This gave support to use a PTAT current to bias the PNP transistors. Later, the effect of I_S and β spread over V_{BE}'s temperature characteristic was described. The effect of β-mismatching between the PNP transistors and their equivalent effect to a β spread was given too.

Finally, after a short comment on PNP transistors in CMOS technologies, the most important application-related requirements used to derive the specifications of the sensor were presented.

CHAPTER 3

Sensor Front-End

3.1 Introduction

In this chapter, the sensor front-end is discussed. Such term (analog) sensor front-end is typically used to describe the (analog) sensor circuitry interfacing with the sensing element (in this work the part of the sensor that ultimately generates the temperature information, e.g. the substrate PNP transistors), whose output (front-end output) is afterward processed by the A/D converter or sensor digital interface. Since in this work sensor front-end and sensing elements are merged into the same circuitry, the term sensor front-end is used indistinctly throughout this text, as follows.

Initially, some concepts for bandgap references are briefly introduced. Some important related signals are thus reviewed, e.g. the proportional to absolute temperature (PTAT) voltage ΔV_{BE}, the complementary to absolute temperature voltage V_{BE} and the temperature-independent or reference voltage V_{REF}, built from the former two voltages. After, the ratiometric temperature measurement is described showing how to obtain a temperature reading by ratioing the PTAT and the reference voltages.

Several important aspects related to the front-end design, e.g. accurate current ratio set up and bias current generation, are extensively analyzed in face of power consumption constraints, as well error cancellation techniques, e.g. dynamic element matching (DEM), off-set cancellation (chopping) and β-compensation. Important accuracy and power consumption trade-offs are then systematically established, which allows to set design parameters given accuracy and power specifications.

After, the front-end implementation is presented where most of the results previously found are employed. Special attention is given to the front-end OTA and current mirror, in terms of their power consumption and accuracy.

Issues related to sensor calibration and trimming and also to V_{BE} curvature, as previously introduced in Chapter 2, are addressed in the end of this chapter.

3.2 A smart temperature sensor

Smart and integrated temperature sensors (using bipolar transistors) are often built using the same back-ground developed for bandgap references. Before going deeply into details concerning the sensor front-end designs, a short overview about bandgap references is given. The ratiometric temperature measurement concept is then explained, and a simplified block diagram of the sensor present.

3.2.1 Concepts for bandgap references

In bandgap references, the reference voltage V_{REF} is obtained by compensating the base-emitter voltage $V_{BE}(T)$ of a bipolar transistor for its temperature dependence [32]. There are many ways of doing this, but essentially a compensating voltage is added to V_{BE} to cancel for at least the first-order temperature dependence of $V_{BE}(T)$. Since $V_{BE}(T)$ decreases with temperature, a compensating voltage proportional to absolute temperature (PTAT) has to be used. Because often the PTAT voltage is small, it has to be amplified prior to the adding, according to what is illustrated in Fig. 3.1.

A substrate PNP transistor Q_L biased through its emitter with a current I_{bias} provides a voltage V_{BE} given by

$$V_{BE}(T) = V_T \ln\left(\frac{I_{bias}}{I_S(T)}\right) \tag{3.1}$$

where $V_T = \frac{k.T}{q}$ is the thermal voltage and I_S is the PNP transistor saturation current.

A voltage $\Delta V_{BE} = V_{BE,R} - V_{BE,L}$ where

$$V_{BE,L}(T) = V_T \ln\left(\frac{I_{bias,L}}{I_{S,L}(T)}\right) \tag{3.2}$$

$$V_{BE,R}(T) = V_T \ln\left(\frac{I_{bias,R}}{I_{S,R}(T)}\right), \tag{3.3}$$

proportional to absolute temperature (PTAT), is obtained by biasing two equal (i.e. same emitter area) PNP transistors Q_L and Q_R at $1 : m$ current ratio, i.e. $m = \frac{I_{bias,R}}{I_{bias,L}}$. Provided mismatching errors in the bias currents and

PNP transistors are eliminated, this voltage is intrinsically accurate [22][23] given by

$$\Delta V_{BE}(T) = V_T \ln(m). \tag{3.4}$$

A temperature-independent reference voltage V_{REF} is obtained by combining V_{BE} with a scaled ΔV_{BE}

$$V_{REF} = \alpha \Delta V_{BE} + V_{BE} \tag{3.5}$$

where α is the scale factor required to compensate for the variation of V_{BE} with the temperature, about -2.2mV/°C, given the variation of ΔV_{BE} with the temperature, e.g. $+0.14$mV/°C if $m = 5$ (see Fig. 3.6(b)). In other words,

$$\frac{\partial V_{BE}}{\partial T} + \alpha \frac{\partial \Delta V_{BE}}{\partial T} = 0 \tag{3.6}$$

which results in

$$\alpha = -\frac{\frac{\partial V_{BE}}{\partial T}}{\frac{\partial \Delta V_{BE}}{\partial T}} = -\frac{\partial V_{BE}}{\partial T} \frac{1}{\frac{k}{q} \ln(m)}. \tag{3.7}$$

For a typical $\frac{\partial V_{BE}}{\partial T} = -2.2$mV/°C, and a current ratio range m from 2 to 10, values of α from 36.8 to 11.1 are found.

3.2.2 Ratiometric temperature measurement

Temperature can be measured by comparing a voltage proportional to absolute temperature, i.e. V_{PTAT}, with a voltage that is immune to temperature, i.e. reference voltage V_{REF}[1]. Because ΔV_{BE} is a PTAT voltage (and it will keep this property when amplified by a constant factor, e.g. the same α as in (3.7) for simplicity), information about the temperature is available if the ratio

$$\mu = \frac{V_{PTAT}}{V_{REF}} = \frac{\alpha \Delta V_{BE}}{V_{BE} + \alpha \Delta V_{BE}} = \frac{\alpha V_T \ln(m)}{V_{BE} + \alpha V_T \ln(m)}, \tag{3.8}$$

which is also PTAT, is evaluated. Because this ratio is dependent on a reference voltage, ratiometric temperature measurement uses the same fundamental concepts for bandgap circuits.

Furthermore, in case of a smart temperature sensor, the same A/D converter used to convert the temperature information to the digital format

[1] In order to increase the temperature sensitivity, the base-emitter voltage V_{BE} may be subtracted from the amplified PTAT voltage [27]. This may relax the resolution requirements for the A/D converter, as latter addressed in Chapter 4.

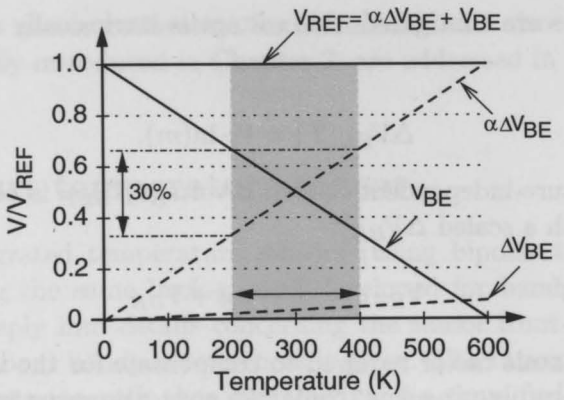


Figure 3.1: Typical voltages V_{BE}, ΔV_{BE}, $\alpha\Delta V_{BE}$ and V_{REF} of a bandgap circuit.

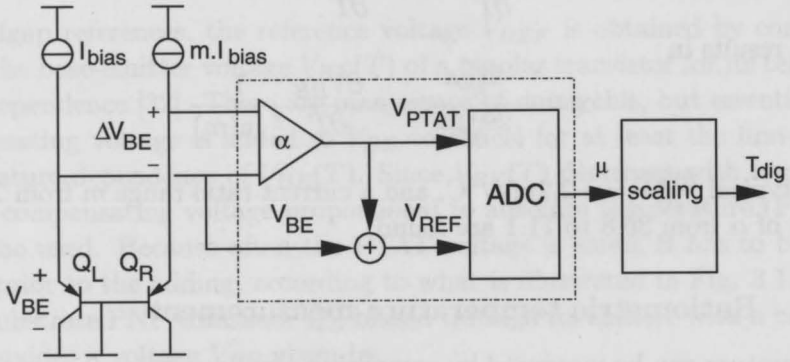


Figure 3.2: Simplified circuit diagram of the smart temperature sensor.

process the two input signals ΔV_{BE} and V_{REF} so to generate the ratio μ, thus providing a temperature reading as

$$T_{dig} = A\mu + B \tag{3.9}$$

where $A \approx 600$ and $B = 0$ are the angular and linear coefficients for a direct result in K. This is schematically shown in Fig. 3.2.

3.3 Accuracy requirements and power constraints

When designing for high accuracy, making all error sources negligible by design is the primary goal. When the design has specified power constraints, it is also fundamental to quantify the amount of power used to make each

error source negligible, restricting it if necessary. Regardless of the case, the first step is to identify the error source, finding how each error affects the temperature measurement. This can be done by calculating the sensitivity of the temperature (as given in (3.9)) to changes in V_{BE}, ΔV_{BE} (or m) and α. The second step is to trade off accuracy and power, setting for instance minimum currents to obtain a given accuracy, or maximum currents to obtain a given power dissipation (at a given supply voltage).

3.3.1 Error and power budgeting

The system implementing the sensor is not ideal, i.e. it suffers from many non-idealities that affect its performance. These non-idealities act by introducing errors to those quantities, e.g. V_{BE}, ΔV_{BE}, α, m, etc. used to build the sensor output. It is important to understand them, so to be able cancel or compensate their effect in the measurement. Since the overall temperature inaccuracy is the combined result of all these errors, often a fraction of the maximum overall temperature error has to be allocated to each of the error sources [22]. For example, in [22], most of the errors were limited down to $\frac{1}{10}$ of the final specified inaccuracy.

However, when designing under other constraints, e.g. power constraints, a general rule limiting all error sources down to one tenth of the total inaccuracy is no longer acceptable, since it may cause power consumption in excess. While a power budgeting can be established in a similar way, e.g. given an overall power consumption, a fraction of this overall value has to be allocated to each component circuit, the problem is better looked as an optimization problem: what is the minimum power consumption given a maximum overall temperature inaccuracy and temperature operating range?

Furthermore, because most of the errors can be solved by means of a larger biasing or of a more complex circuit, the error budgeting ends up subjected to the power budgeting. In doing this, determining how an error introduced, for instance, in ΔV_{BE} is translated to the output (in °C) is extremely important.

3.3.2 Temperature reading errors - Errors in V_{BE}, ΔV_{BE}, α and m in the temperature reading

The temperature reading accuracy and the power required to obtain it are the most important sensor's performance figures in this work. The sensitivities (partial derivatives) of the temperature as given by (3.9) with respect to V_{BE},

ΔV_{BE}, α [22] and m

$$\frac{\partial T_{dig}}{\partial V_{BE}} = A\frac{-\alpha\Delta V_{BE}}{(V_{BE}+\alpha\Delta V_{BE})^2} = -\frac{T}{V_{REF}} \tag{3.10}$$

$$\frac{\partial T_{dig}}{\partial \Delta V_{BE}} = A\frac{\alpha V_{BE}}{(V_{BE}+\alpha\Delta V_{BE})^2} = \frac{(A-T)}{V_{REF}}\alpha \tag{3.11}$$

$$\frac{\partial T_{dig}}{\partial \alpha} = A\frac{\Delta V_{BE} V_{BE}}{(V_{BE}+\alpha\Delta V_{BE})^2} = \frac{T}{\alpha}(1-\frac{T}{A}) \tag{3.12}$$

$$\frac{\partial T_{dig}}{\partial m} = A\frac{\alpha V_T V_{BE}}{m(V_{BE}+\alpha\Delta V_{BE})^2} = \frac{\alpha}{mV_{REF}} V_T A\left(1-\frac{T}{A}\right) \tag{3.13}$$

provide important relations (between each parameter variation and the sensor's output variation) that help the sensors design. With the sensitivity, the calculation of the maximum error or deviation allowed in each parameter specified a maximum error contribution in degree Celsius (°C) is straightforward.

However, because the sensitivities depend on T, often they have to be evaluated at temperatures where the worst cases appear. It is not difficult to see from equations (3.10-3.13) that the worst case scenario for V_{BE} is at high temperature ($T \approx 400$K), while for ΔV_{BE} at low temperature ($T \approx 200$K) and for α and m in the middle range ($T = 300$K). So taking this into consideration, the following error limits are established for V_{BE}, ΔV_{BE}, α [22] and m given an error contribution ε_T in the temperature:

$$|V_{BE} - V_{BE,ideal}| = -\frac{V_{REF}}{T}\varepsilon_T < \frac{3\text{mV}}{^\circ\text{C}}\varepsilon_T \tag{3.14}$$

$$|\Delta V_{BE} - \Delta V_{BE,ideal}| = \frac{V_{REF}}{(A-T)\alpha}\varepsilon_T < \frac{3\text{mV}}{^\circ\text{C}}\frac{1}{\alpha}\varepsilon_T \tag{3.15}$$

$$\left|\frac{\alpha - \alpha_{ideal}}{\alpha_{ideal}}\right| = \frac{A}{(A-T)T}\varepsilon_T < \frac{2}{3}\%\frac{1}{^\circ\text{C}}\varepsilon_T \tag{3.16}$$

$$\left|\frac{m - m_{ideal}}{m_{ideal}}\right| = \frac{V_{REF}}{(A-T)V_T}\frac{1}{\alpha}\varepsilon_T < \frac{15.5}{\alpha}\%\frac{1}{^\circ\text{C}}\varepsilon_T \tag{3.17}$$

where $V_{REF} \approx 1.2$V and $A \approx 600$.

In this way, given a maximum error contribution, e.g. $\varepsilon_T = 0.01$°C, a worst case variation of 30μV in V_{BE}, 1.8μV in ΔV_{BE} (for $\alpha = 16$), 0.0067% in α and 0.01% (for $\alpha = 16$, $T = 300$K) in m is allowed. Such limits set fundamental specifications that help in the sensor design, thus being used throughout this work[2].

[2]While the sensitivity of the temperature with respect to α is discussed here, results are only used in the next chapter.

3.4 The current ratio m

In this section, most of the aspects related to the current ratio m are addressed. One of these aspects is its accuracy. Since the PTAT voltage ΔV_{BE} depends directly on this ratio, it is important to design it accurately. A second aspect is the power consumption required to achieve the given accuracy, which ultimately establishes important trade-offs.

3.4.1 Designing an accurate current ratio m

The current ratio m is the only actual design parameter left in the voltage ΔV_{BE}, as given in (3.4). In this way, the full accuracy of ΔV_{BE} will depend on that of the current ratio m (indeed the current-density ratio when the PNPs have the same emitter areas). That is why is so important to design it very accurately. Its value, nonetheless, has to be established taking both accuracy and power consumption requirements into account.

3.4.1.1 Error due to mismatch in the current sources

A set of $(1+m)$ CMOS current sources can be used to generate a 1 to m current ratio as shown in Fig. 3.3. If the current sources are ideal, the current ratio m is also ideal and the voltage ΔV_{BE} is accurately established. In practice, however, current sources suffers from mismatch. The most common reasons of such mismatch are due to threshold ΔV_{th} and current factor $\Delta\beta$ mismatches (refer to Subsection 3.8.1).

When a larger current needs to be implemented, a parallel combination of m identical copies of the unit current source should be used [35]. Even in this case, the resultant current is not exactly m times the unit one, but m plus an error, i.e. $m+\Delta m$, which results in a voltage ΔV_{BE} given by

$$\Delta V_{BE} = V_T \ln(m+\Delta m) \tag{3.18}$$

$$\approx \Delta V_{BE}|_{\Delta m=0} + \frac{kT}{q}\frac{\Delta m}{m} \qquad (\Delta m \ll m). \tag{3.19}$$

The second term in (3.19) is an approximation to the absolute error introduced in ΔV_{BE} due to an error Δm in the current ratio m. Assuming that Δm is weakly dependent on m, the larger the current ratio is, the smaller the ratio $\frac{\Delta m}{m}$ and the error due to mismatch are. Moreover, assuming $\frac{\Delta m}{m}$ constant with temperature, this type of error is expected to be PTAT.

Assuming, for example, a mismatch $\frac{\Delta m}{m} = 0.1\%$, the absolute error in ΔV_{BE} is about 26μV at room temperature. If $m = 5$, then $\alpha = 16$ and the sensitivity as given in (3.11) is 4K/mV. Therefore, the temperature error

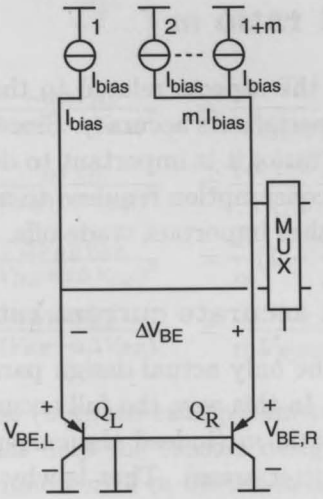


Figure 3.3: Front-end schematic (core) showing $(1+m)$ current sources used to bias the PNP transistors with a 1 to m current ratio.

due to mismatch amounts to 0.1K. If the final inaccuracy is expected to be below 0.1°C, an error of 0.1°C (100% of the overall temperature error) only due to one source of error is not acceptable. This, added to the fact that the assumed mismatch can be worse than 0.1%, is a clear indication that the current matching level has to be improved by some technique.

Evidently, the current ratio can be designed larger. For $m = 10$, then $\alpha \approx 10$ and the sensitivity is now 2.5K/mV. The temperature error due to mismatch in this case is then 0.065K, roughly 35% less than the former case, but at the price of a higher power consumption (given the same unit current source)[3].

3.4.1.2 Error due to mismatch in the current sources with Dynamic Element Matching

Dynamic element matching (DEM) is a well-known technique [36][37] used to improve the effective accuracy of a system. Such technique is helpful when there are several nearly equal circuit elements that can be dynamically and/or cyclically interchanged (position and/or function) over time, so mismatching errors are averaged out. The effective accuracy is often some orders of magnitude higher the relative accuracy of such elements themselves [15].

[3]Even in this case, the error might be considered too large, taking into account a final inaccuracy below 0.1°C.

Figure 3.4 shows how DEM of the bias current sources can be implemented [38][27][39]. Every ΔV_{BE} cycle, a bias current $I_{bias,j}$ $(1 \leq j \leq m+1)$ is used as a unit current source to bias the left PNP transistor, while the remaining current sources are directed to the right one, resulting in a ΔV_{BE} given by

$$\Delta V_{BE,j} = \frac{kT}{q} \ln \left(\frac{\sum_{i=1, i \neq j}^{m+1} I_{bias,i}}{I_{bias,j}} \right) = \frac{kT}{q} \ln \left(m + \Delta m_j \right), \; 1 \leq j \leq m+1, \tag{3.20}$$

which turns, after $m+1$ ΔV_{BE} cycles, into an averaged $\Delta V_{BE,avg}$ given by

$$\Delta V_{BE,avg} = \frac{1}{m+1} \sum_{j=1}^{m+1} \Delta V_{BE,j}. \tag{3.21}$$

It can be shown that first-order errors are canceled due to averaging, while second-order errors[4] remain given by [22]:

$$|\Delta V_{BE,avg} - \Delta V_{BE}|_{\Delta m=0}| < \frac{1}{2} \frac{kT}{q} \left(\frac{\Delta m}{m} \right)^2, \tag{3.22}$$

where $\frac{\Delta m}{m}$ is the worst-case mismatch between the currents.

In this case, even for a much worse mismatching assumption ($\frac{\Delta m}{m} = 1\%$), the absolute error in ΔV_{BE} is about 1.3μV at room temperature. For the same sensitivity of 4K/mV at $T = 300$K, the temperature error due to mismatch amounts to 0.005K. This result shows the effectiveness of DEM.

3.4.1.3 Error due to finite output impedance

Another error in the current ratio m (which occurs even for matched current sources) is due to the finite output impedance of the $(1+m)$ current sources used to bias the PNP transistors. This is illustrated in Fig. 3.5 where the finite output impedance is modeled as r_{DS}. In the left side, the ideal current I is added of $V_{DS,L}/r_{DS}$ where $V_{DS,L} = V_{supply} - V_{BE,L}$, while in the right side, every ideal current I is added of $V_{DS,R}/r_{DS}$ where $V_{DS,R} = V_{supply} - V_{BE,R}$. Because the voltage $V_{DS,L}$ and $V_{DS,R}$ are different (due to the established $\Delta V_{BE} = V_{BE,R} - V_{BE,L}$), the current ratio m is now given by

$$\begin{aligned} \frac{I_{bias,R}}{I_{bias,L}} &= \frac{m\left(I + \frac{V_{supply} - V_{BE.R}}{r_{DS}}\right)}{I + \frac{V_{supply} - V_{BE,L}}{r_{DS}}} \\ &= m\left(1 - \frac{\Delta V_{BE}}{I_{bias,L} r_{DS}}\right). \end{aligned} \tag{3.23}$$

[4] Third and higher-order errors were neglected.

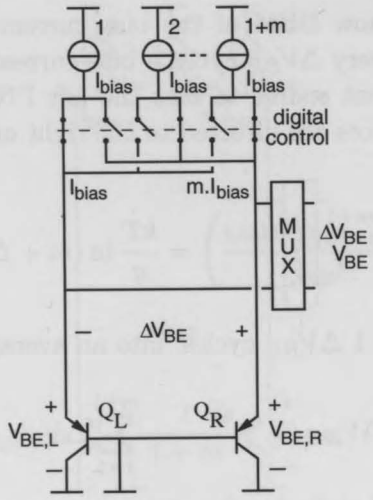


Figure 3.4: Front-end schematic (core) showing $(1+m)$ current sources used to bias the PNP transistors with a 1 to m current ratio. Switches (digitally controlled) are used to dynamically interchange the current sources, so to implement DEM.

The term inside the parenthesis in (3.23) that appears subtracting the unity is the relative current ratio error, i.e. $\frac{\Delta m}{m} = -\frac{\Delta V_{BE}}{I_{bias,L} r_{DS}}$.

In order to reduce the error, either the voltage ΔV_{BE} has to be reduced or the product $I_{bias,L} r_{DS}$ increased. A reduction of ΔV_{BE}, however, is counter balanced by the requirement of an increased α, which in turn hardens the accuracy requirements for $\frac{\Delta m}{m}$ (according to (3.17)). Additionally, an increase in $I_{bias,L}$ is not accepted since it will affect directly the power consumption in the front-end.

The option left is to increase r_{DS}, a similar result found previously in [22] where cascoded current sources were used, diminishing the effect of the finite output conductance in the current ratio m. Since cascoding does not turn into additional bias requirements, it was adopted in this work too, despite of a poorer signal swing at the current source output.

3.4.2 The current ratio m value

Dynamic element matching can be used to average the current sources mismatching errors out. Cascoding can prevent or limit errors due to the finite current sources output impedance. However, there is still an opened question about the value of m. Looking to (3.22), it seems clear that the larger the

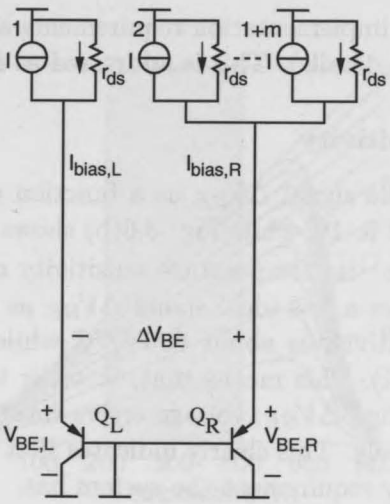


Figure 3.5: Front-end schematic used to model the inaccuracy introduced in the current ratio m due to the finite output impedance r_{DS} of the current sources.

current ratio is, the smaller is the mismatching residual error.

The explanation is based on statistics: the more elements available to calculate the average, the more reliable the average result is (and thus less variation from sensor to sensor). It is worth to mention that, since it is the smallest element (the unit current in the 1 to m current ratio) that determines the accuracy of the ratio, it is the time average of unit current over m cycles that ultimately sets the ratio accuracy. Evidently, given a unit current, the more elements are used, the higher the power consumption is.

Other factors indeed might be relevant. For instance, given two similar transistors, if m is large, the current density in one transistor is relatively large compared to the other, which might cause the transistors to operate under too much different conditions.

Additionally, the current gain β has to be constant under very different biasing [40]. Although the current gain β can be assumed relatively independent of the bias current for some decades [15][22], this is not true in deep submicron technologies. Regardless of case, in general a better β-matching is expected if the two PNPs transistors operate under more similar current densities.

Finally, there is perhaps a more relevant point than the ones just mentioned: the value of m sets the ΔV_{BE} sensitivity with temperature (and its magnitude at a given temperature) and the required amplification factor α

and the accuracy and implementation requirements associated with it (refer to Chapter 4 for more details). This is addressed as follows.

3.4.2.1 ΔV_{BE} Sensitivity

Figure 3.6(a) shows the signal ΔV_{BE} as a function of T, as given in (3.4), for values of m from 2 to 10, while Fig. 3.6(b) shows the variation of ΔV_{BE} as a function of T, i.e. its temperature sensitivity or $\frac{\partial(\Delta V_{BE})}{\partial T}$, for each m (Fig. 3.6(a) also shows a less ideal signal ΔV_{BE} as latter given in (3.25)). For $m = 2$, the sensitivity is about 60μV/°C while for $m = 10$, close to 200μV/°C ($T \approx 300$K). This means that, in order to prevent errors larger than 0.1°C when reading ΔV_{BE}, voltage errors smaller than 6μV and 20μV are required, respectively. This clearly indicates that the larger the m is, the less stringent accuracy requirement the system has.

Alternatively, the same result can be obtained by addressing the inter-dependence between the current ratio m and the amplification factor α (Table 3.1 shows the required amplification factor α for a current ratio m varying from 2 to 10, given a typical $\frac{\partial V_{BE}}{\partial T} = -2.2$mV/°C, as introduced in Section 3.2.1). The larger the m (and ΔV_{BE}), the smaller the amplification factor α needed to properly amplify ΔV_{BE} so to compensate the variation of V_{BE} with temperature. Such smaller α decreases accuracy requirements for the A/D converter (by reducing accuracy requirements when reading ΔV_{BE}, as given in (3.15)).

However, Fig. 3.6(b) shows that the increase in m does not cause the same increase in the sensitivity as when m is small[5]. While the relative increase in both sensitivity and number of current sources are approximately linearly dependent, there is no sense in increasing m without response in the sensitivity. This gives an indication of a maximum recommended value for m if power constraints are set.

3.4.2.2 Error due to the saturation current and switch leakage

When the PNPs bias current is set to, e.g a few hundreds of nA, aiming less power consumption than in the prior art [22] while keeping $(1+m)$ elements for DEM, errors in the current ratio m due to the saturation current I_S and switch leakage might be no longer negligible. Because these currents add to or subtract from the main bias current, they may introduce significant error in the current ratio.

For instance, if the saturation current I_S has a value relatively comparable with the bias current, the value of V_{BE} given in (3.1) is not more accurate.

[5]This is due to the logarithmic function in the sensitivity, i.e. $\frac{\partial(\Delta V_{BE})}{\partial T} = \frac{k}{q}\ln(m)$.

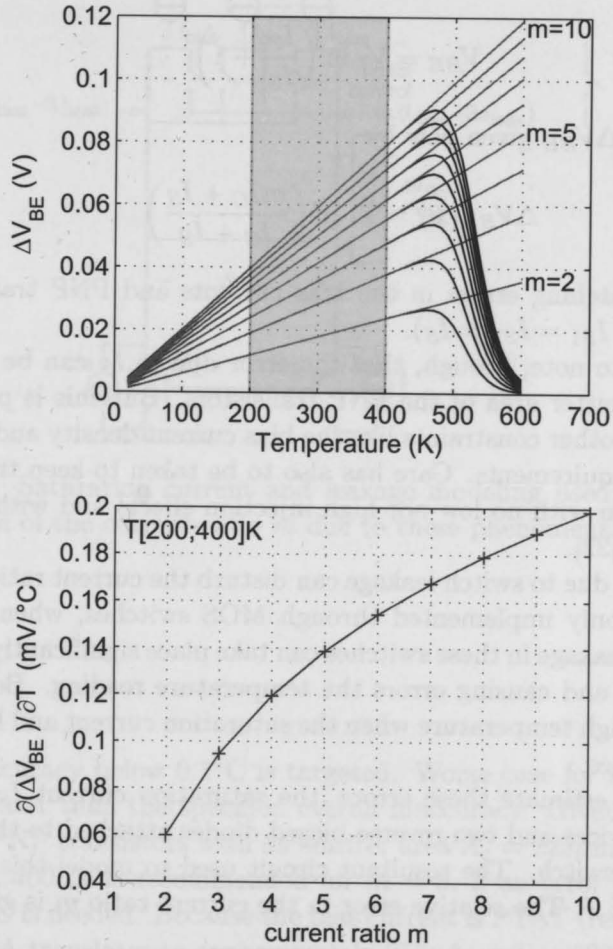


Figure 3.6: Simulated (a) ΔV_{BE} (as given in (3.4) and (3.25)) as a function of T, from 200K to 400K (i.e. approximately from -55°C to 125°C), for values of m from 2 to 10, and (b) $\frac{\partial(\Delta V_{BE})}{\partial T}$ as a function of m.

Table 3.1: Approximate amplification factor α for different current ratios m.

	current ratio m								
	2	3	4	5	6	7	8	9	10
α	36.8	23.2	18.4	15.6	14.2	13.1	12.3	11.6	11.1

A better approximation for V_{BE} is given by

$$V_{BE} = V_T \ln\left(\frac{I_C}{I_S} + 1\right) \tag{3.24}$$

which means a ΔV_{BE} given now by

$$\Delta V_{BE}(T) = V_T \ln\left(\frac{mI_C + I_S}{I_C + I_S}\right), \tag{3.25}$$

provided mismatching errors in the bias currents and PNP transistors are eliminated (i.e. $I_{S1} = I_{S2} = I_S$).

Interesting to note, though, that the error due to I_S can be reduced by reducing the emitter area of the PNP transistors. But this is prevented to some extent by other constraints like the bias current density and by transistor matching requirements. Care has also to be taken to keep transistors in a working region with no low nor high injection effects and with a constant current gain β [22].

The current due to switch leakage can disturb the current ratio too. Since DEM is commonly implemented through MOS switches, when the bias is relatively low, leakage in these switches can take place significantly, modifying the value of m and causing errors the temperature reading. Both case are even worse at high temperature when the saturation current and leakage have the largest values.

In order to estimate these errors, the saturation current I_S was added to the PNP devices and two reverse-biased diodes attached to the drain and source of each switch. The resultant circuit used to model this situation is shown in Fig. 3.7. The relative error in the current ratio m is given by

$$\frac{\partial m}{m} = \frac{\frac{mI - [(1+1)I_{leak,max}] + I_S}{I - [(1+m)I_{leak,max}] + I_S} - \frac{mI_{bias}}{I_{bias}}}{\frac{mI_{bias}}{I_{bias}}} \tag{3.26}$$

where $I = I_{bias} - 3.I_{leak,max}$ and $I_{S,max}$ and $I_{leak,max}$ are the worst (i.e. highest temperature, 140°C) case saturation current and diode leakage for a minimum size NMOS switch [28], respectively.

Simulation results of the error due to the saturation current and leakage obtained from (3.26) are plotted in Fig. 3.8 for an estimated $I_{leak,max} =$ 1.71pA, $I_{S,max} = 110$pA (PNP, $A_e = 225\mu m^2$, 140°C) and I_{bias} from 100nA to 1μA. Effects due to I_S and leakage do not add together. The error is mainly due to I_S though, since for the PNP actual sizes $I_{S,max} \gg I_{leak}$. The trade-off between accuracy and biasing is very clear. A bias current of 100nA, even for $m = 10$, gives an error close to 0.05°C, which is too large if

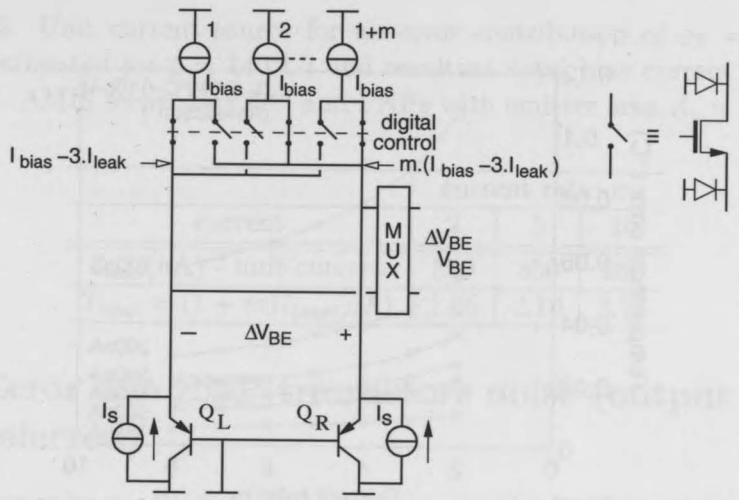


Figure 3.7: Saturation current and leakage modeling used to estimate the degradation of the current ratio m due to these phenomena.

a final inaccuracy below 0.1°C is targeted. Worse case for $m = 2$ when the error is greater than the specified overall inaccuracy. Given the technology ([28]) and PNP transistors with an emitter area $A_e = 225\mu m^2$, a bias current larger than 400nA is recommended for $m = 5$, if an error not higher than $\varepsilon_T = 0.02°C$ is needed. Because the bias current is PTAT (refer to subsection 3.6.1) 400nA translates to approximately 250nA at $T = 300K$.

Another way of viewing the same situation is shown in Fig. 3.9, where the temperature error is plot against the bias current instead of the current ratio m. For a given m, the error contribution decreases when the bias current is increased. In order to keep the error below a given level, the bias current has to be increased sufficiently. The smaller the current ratio m is, the larger is the bias current required for the same error contribution.

It is worth mentioning, however, not only the unit bias current but the total bias $I_{total} = (1+m)I_{bias}$. For instance, the smaller the current ratio m is, the larger is the required unit bias current for a given error contribution, but the smaller the total current. This situation is summarized in Table 3.2 for an error contribution of $\varepsilon_T = 0.02°C$. Despite of the poorest sensitivity of ΔV_{BE}, a current ratio $m = 2$ is the best alternative if only power consumption is being taken into account.

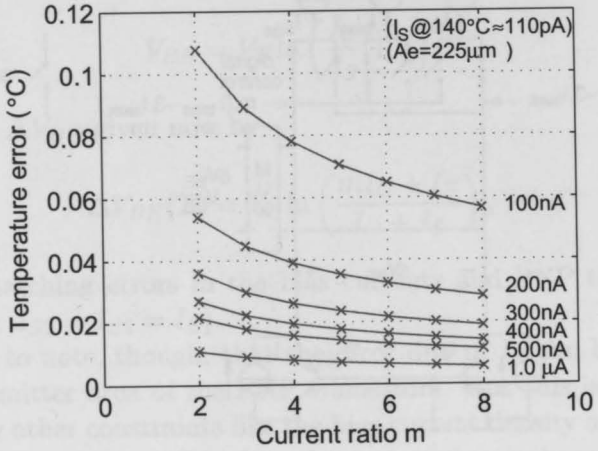


Figure 3.8: Temperature error for as a function of the current ratio m, for I_{bias} from 100nA to 1μA. Results obtained for $I_{S,max}$ = 110pA (estimated for T = 140°C) and $I_{leak,max}$ = 1.71pA (AMIS technology [28] and PNPs with emitter area $A_e = 225\mu m^2$).

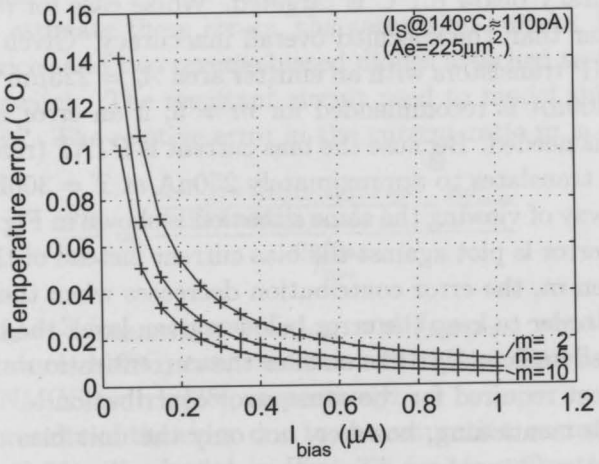


Figure 3.9: Temperature error as a function of the bias current I_{bias}, for m = 2, 5 and 10. Results obtained for the same conditions mentioned in Fig. 3.8.

Table 3.2: Unit current source for an error contribution of $\varepsilon_T = 0.02°C$ (values estimated for $t = 140°C$) and resultant total bias current for each situation. AMIS technology [28] and PNPs with emitter area $A_e = 225\mu m^2$.

	current ratio m		
current	2	5	10
I_{bias}(nA) - unit current	550	350	250
$I_{total} = (1+m)I_{bias}(\mu A)$	1.65	2.10	2.75

3.5 Error due PNP transistors noise (output referred)

In this particular application, the noise generated by PNP transistors is negligible. According to [22], the error introduced by the noise of the bipolar front-end, which can be applied to this work too, is given by

$$\sigma_T = \frac{A}{V_{REF}}\alpha\sqrt{\frac{2kT}{g_m}\frac{(1-\mu)}{Nt_{clk}}} \tag{3.27}$$

where $g_m = \frac{I_{bias}}{V_T}$ is the transconductance of the bipolar transistor and N is the number of $\Delta\Sigma$ cycles (as defined in the next chapter). The error introduced is inversely proportional to $\sqrt{I_{bias}}$, and it is of the order of mK, even if the bias current is reduced to levels below 100nA. e.g. given $A \approx 600$ and $V_{REF} = 1.2V$, the error at room temperature ($\mu = 0.5$) for $\alpha = 16$, $N = 100$, $t_{clk} = \frac{1}{10}$kHz and $I_{bias} = 100$nA is 2.6mK. Therefore, the noise that comes from the PNP transistors is not a limiting factor for bias reduction, since the output noise will be mainly determined by noise introduced by the read out circuitry as studied in Chapter 4.

3.6 The bias current

The voltage ΔV_{BE} is in theory not dependent on the bias current amplitude, but only on the current-density ratio instead. The voltage V_{BE}, on the other hand, is greatly dependent on it (as addressed in Chapter 2). In order to minimize design variations of the bias current (designing it very accurately) so to generate a more accurate V_{BE}, reducing spread from sensor to sensor, special circuitry has been used in help of the front-end core. Obviously, any auxiliary circuit ends adding to the system's power consumption. It is important to design it so to achieve the accuracy benefit at low additional power.

3.6.1 Designing an accurate bias current

Many topologies have been considered to generate accurately a bias current [2][41][22]. The types are: complementary to the absolute temperature (CTAT), constant with temperature (TI) and proportional to absolute temperature (PTAT). The first one often uses the voltage V_{BE} of a single bipolar transistor to generate I_{bias} resulting in

$$I_{bias} = \frac{V_{BE,bias}}{R_{bias}}. \tag{3.28}$$

Because V_{BE} decreases with temperature, I_{bias} decreases too. Assuming R_{bias} temperature independent, the bias current is CTAT. Obviously, because V_{BE} depends on the saturation current I_S of the PNP transistor, the bias current suffers strongly from process spread, since both I_S and R_{bias} spread with the fabrication process.

Biasing with a constant current has also been investigated [42]. In this case, a PTAT and a CTAT currents are added together, so to generate a first order temperature independent bias current. By biasing a resistor with this current, instead of a PNP transistor, a 1V supply-voltage bandgap can be built. However, this type of biasing suffers strongly from process spread too and the circuit is relatively more complex.

The third type is the PTAT biasing. Despite of the former biasing topologies, a PTAT bias current given by

$$I_{bias} = \frac{\Delta V_{BE,bias}}{R_{bias}} = V_T \ln(p) \frac{1}{R_{bias}} \tag{3.29}$$

has been preferred [23]. This type of temperature dependency has proved to be very useful helping to reduce the V_{BE} curvature (refer to Section 3.10). Moreover, and perhaps more importantly, it has a lower spread since it does not depend on V_{BE} (I_S) but only on R_{bias} and current ratio p. This strongly enhances the initial bias accuracy, which may ultimately contribute to reduce the non-trimmed sensor's inaccuracy.

A typical circuit that provides a bias current with such characteristic is shown in Fig. 3.10 [41][22]. The circuit, referred as pre-bias circuit, uses the same classical approach for bandgap circuits to generate the bias current. A set of $(1 + p)$ PMOS current sources is used to bias two substrate PNP transistors Q_1 and Q_2. A feedback loop helped by an OTA enforces a voltage $\Delta V_{BE,bias}$ over the resistor R_{bias}. The result is a PTAT bias current I_{bias} that depends on the current ratio p and on the bias resistor R_{bias}.

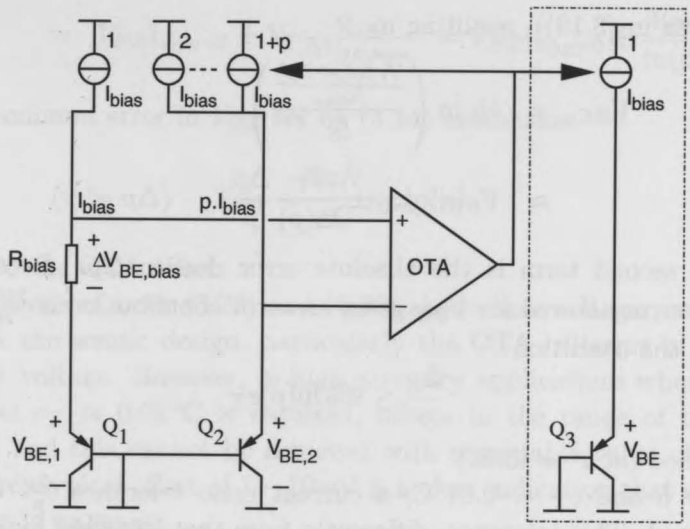


Figure 3.10: Pre-bias circuit used to generate a PTAT bias current I_{bias}, mirrored to the core (schematically represented at the right side).

3.6.2 The bias current and the current ratio p of the bias circuit

The power dissipated in the pre-bias circuit is directly related to the amplitude of the unit current source used to bias the PNP transistors and to the current ratio p since it sets the total number $(1+p)$ of current sources used.

Because of the architecture very similar to the core, i.e. two PNPs Q_1 and Q_2 are biased at 1 to p current ratio, the same analysis used for the core can be used here to set minimum bias current and current ratio p given accuracy requirements. However, since the pre-bias is an auxiliary circuit, the way the errors in this circuit affect the sensor accuracy is different, so other specifications are expected.

Using the bias current generated in the pre-bias circuit, as given in (3.29), to bias the PNP transistors in the core, the following V_{BE}

$$V_{BE} = V_T \ln\left(\frac{\frac{\Delta V_{BE,bias}}{R_{bias}}}{I_S}\right) = V_T \ln\left(\frac{\frac{V_T \ln(p)}{R_{bias}}}{I_S}\right) \tag{3.30}$$

is obtained. Because the circuit is not ideal, V_{BE} will deviate from its ideal value. The spread of the resistor R_{bias} and of the saturation current I_S are partially responsible for such deviation. The error Δp in the current ratio p also contributes to this deviation since it modifies $\Delta V_{BE,bias}$ (in the same

way as given in (3.19)), resulting in

$$V_{BE} = V_T \ln\left(\frac{\frac{V_T \ln(p+\Delta p)}{R_{bias}}}{I_S}\right) \tag{3.31}$$

$$\approx V_{BE}|_{\Delta p=0} \pm \frac{V_T}{\ln(p)}\frac{\Delta p}{p} \qquad (\Delta p \ll p) \tag{3.32}$$

where the second term is the absolute error due to Δp. Because of the maximum error allowed for V_{BE} given an error contribution is $\frac{3\text{mV}}{^\circ\text{C}}\varepsilon_T$ (refer to (3.14)), the condition

$$\frac{\Delta p}{p} \leq 9\% \ln(p)\varepsilon_T \tag{3.33}$$

is established (at $T = 400\text{K}$).

If $p = 5$ and $\varepsilon_T = 0.01^\circ\text{C}$, a current ratio tolerance $\frac{\Delta p}{p} < 0.15\%$ is then required. This tolerance, differently from that specified for the core, as given in (3.17), can be more easily established (and without DEM of current sources). However, the saturation current I_S of the PNP transistors used in the pre-bias circuit can modify the current ratio p in the same way it modifies the current ratio m in the core, as discussed in Section 3.4.2.2.

For PNP transistors of same emitter area of those in the core, i.e. $A_e = 225\mu\text{m}^2$, a bias current in the pre-bias circuit greater than 100nA (as can be calculated from (3.26)) already meets the requirement given in (3.33) so to prevent an error in the current ratio p due to the saturation current and leakage. A bias current of 125nA, half of the bias current in the core (at room temperature) was chosen. This current is important in the sense it reveals that the bias current in the pre-bias can be set smaller in regard to that in the core, so an accurate biasing can be generated under relatively lower power.

However, in order to have PNP transistors of both core and pre-bias working with the same current density, PNPs in the pre-bias circuit were thus designed having consequently half of the area of PNPs in the core, i.e. $A_e = 112\mu\text{m}^2$. This assures that no error due to saturation current is expected in the pre-bias circuit due to a smaller saturation current now[6].

The OTA's offset V_{OS} is another source of error. Because it adds directly to $\Delta V_{BE,bias}$, it modifies the bias current affecting consequently V_{BE}:

$$V_{BE} = V_T \ln\left(\frac{\frac{\Delta V_{BE,bias} \pm V_{OS}}{R_{bias}}}{I_S}\right) \tag{3.34}$$

[6]This triggers the possibility for a further reduction in the bias current and again a new area reduction for sake of the same current density. At some point, this is prevented by mismatching requirements.

$$\approx V_{BE}|_{V_{OS}=0} \pm V_T \frac{V_{OS}}{\Delta V_{BE,bias}} = V_{BE}|_{V_{OS}=0} \pm \frac{V_{OS}}{\ln(p)} \quad (3.35)$$

The maximum error in V_{BE} set by (3.14) establishes

$$V_{OS} < \frac{3\text{mV}}{^\circ\text{C}} \varepsilon_T \ln(p). \quad (3.36)$$

as a limit to the OTA offset.

The analysis of both (3.33) and (3.36) shows how a larger current ratio p can relax the sensor design, particularly the OTA in terms of maximum input offset voltage. However, in high accuracy applications where an error contribution $\varepsilon_T \approx 0.01^\circ\text{C}$ is required, offsets in the range of tens of μV are needed and this cannot be achieved with reasonable value of p (< 10). Furthermore, typical offset of $1 - 10$mV is a clear indication that some offset cancellation is necessary.

The finite open-loop gain A_{OL} is another source of error. Because of its finite value, the differential voltage at the OTA input is approximately equal to $\Delta V_{BE,bias}/A_{OL}$ instead of zero. It adds to $\Delta V_{BE,bias}$ like the offset does:

$$V_{BE} = V_T \ln \left(\frac{\frac{\Delta V_{BE,bias} \pm \frac{\Delta V_{BE,bias}}{A_{OL}}}{R_{bias}}}{I_S} \right) \quad (3.37)$$

$$\approx V_{BE}|_{A_{OL}\to\infty} \pm V_T \frac{1}{A_{OL}} \quad (1 \ll A_{OL}). \quad (3.38)$$

The same error limit for V_{BE} establishes then

$$A_{OL} > \frac{V_T}{3\text{mV} \cdot \varepsilon_T} \quad (3.39)$$

which means an open-loop gain of about 1150 or 62dB (at $T = 400$K).

3.6.3 Offset cancellation - chopping

In order to generate a bias current that is not affected by the OTA's offset, some offset cancellation technique is required. The most known techniques are auto-zeroing and chopping [43]. Both are well-known techniques and present a wide applicability, e.g. low-power temperature sensors [44]. The chopping technique as described in [43] is implemented by switching the OTA's input and output, as shown in Fig. 3.11, which fits properly the pre-bias circuit architecture, i.e. it runs continuously and has no large capacitors which could be used for offset sampling, as required in auto-zeroing [43].

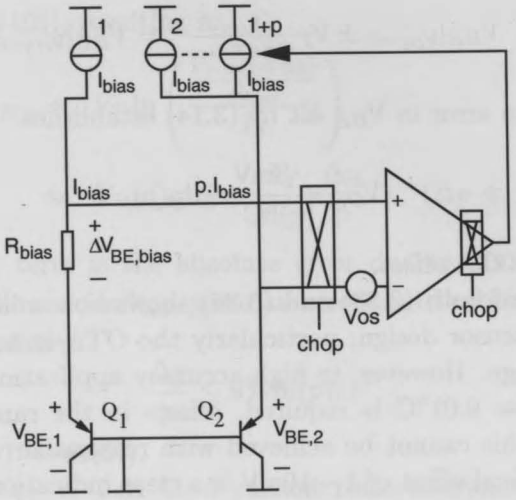


Figure 3.11: Chopping scheme used to compensate the OTA offset.

When the chop signal is 0, the voltage over the bias resistor is

$$V_{R,chop=0} = V_{BE,2} - V_{BE,1} - V_{OS} = \Delta V_{BE,bias} - V_{OS} \tag{3.40}$$

while, when the chop signal is 1, is

$$V_{R,chop=1} = V_{BE,2} - V_{BE,1} + V_{OS} = \Delta V_{BE,bias} + V_{OS}. \tag{3.41}$$

After two consecutive cycles, the average voltage over the bias resistor is then $\Delta V_{BE,bias}$. The main drawback of such technique is the ripple it causes at the OTA's output. The lower the current ratio p is (and the voltage $\Delta V_{BE,bias}$), the higher the ripple and the residual offset after chopping are [24]. Table 3.3 summarizes the expected ripple due to chopping for some values of p (with $\Delta V_{BE,bias}$ calculated at room temperature). In order to reduce the total number of current sources and the power dissipated in the pre-bias (given a unit current), a small p is necessary. However, this implies a small $\Delta V_{BE,bias}$ and consequently a large offset residual that will appear over the generated bias current. While this offset may not harm the generation of ΔV_{BE} in the core, it will cause spread to V_{BE}. In the latter case, a poorer non trimmed sensor performance is expected.

Interesting to mention that with such offset cancellation scheme, while $\Delta V_{BE,bias}$ is free from the offset, it still suffers from PNP transistors mismatch, that causes the approximation:

$$\Delta V_{BE,bias} = V_{BE,2} - V_{BE,1} = V_T \ln\left(\frac{pI_{bias}}{I_{S2}}\frac{I_{S1}}{I_{bias}}\right) \approx V_T \ln(p). \tag{3.42}$$

Table 3.3: Ripple amplitude (%) due to chopping versus the current ratio p (an offset voltage $V_{OS} = \pm 5\text{mV}$ is assumed).

p	ΔV_{BE} @30°C	ripple (%)
2	17.9	±28
5	41.6	±12
10	59.5	±8

One solution to this issue was proposed in [22]. By adding a second bias resistor $R_{1b} = R_{1a} = R_{bias}$, placing the OTA input chopper at the current sources so to bias the transistors at $1 : p$ and $p : 1$ (while keeping a multiplexer in place of the original chopper), both offset cancellation and transistor mismatching are properly handled. This is shown in Fig. 3.12. When the chop signal is 0, the voltage over the bias resistor R_{1a} is

$$V_{R,chop=0} = V_{BE,2a} - V_{BE,1a} - V_{OS} = \Delta V_{BE,bias}(1:p) - V_{OS} \tag{3.43}$$

while when the chop signal is 1, the voltage over the bias resistor R_{1b} is

$$V_{R,chop=1} = V_{BE,2b} - V_{BE,1b} + V_{OS} = \Delta V_{BE,bias}(p:1) + V_{OS}. \tag{3.44}$$

After two consecutive cycles, the average voltage $\Delta V_{BE,bias}$ over the bias resistors is

$$\begin{aligned} \Delta V_{BE,bias} &= \frac{1}{2}\left[\Delta V_{BE,bias}(1:p) - V_{OS} + \Delta V_{BE,bias}(p:1) + V_{OS}\right] \\ &= \frac{1}{2} V_T \left[\ln\left(\frac{pI_{bias}}{I_{S2}}\frac{I_{S1}}{I_{bias}}\right) + \ln\left(\frac{pI_{bias}}{I_{S1}}\frac{I_{S2}}{I_{bias}}\right)\right] \\ &= V_T \ln(p) \end{aligned} \tag{3.45}$$

for an ideal offset cancellation. In this way, the PNPs mismatching issue was solved. With such arrangement, the bias current suffers even less from process spread.

It is, however, the averaged value of V_{BE} (whose average is performed by the $\Delta\Sigma$ converter [22])

$$\frac{1}{2}\left[V_{BE,L} + V_{BE.R}\right] = \frac{1}{2} V_T \left[\ln\left(\frac{I_{bias,L}}{I_{SL}}\right) + \ln\left(\frac{I_{bias,R}}{I_{SR}}\right)\right] \tag{3.46}$$

$$\approx V_T \ln\left(\frac{\Delta V_{BE,bias}}{\sqrt{R_{1a}R_{1b}}\sqrt{I_{SL}I_{SR}}}\right) + -\frac{1}{2} V_T \left(\frac{V_{OS}}{\Delta V_{BE,bias}}\right)^2. \tag{3.47}$$

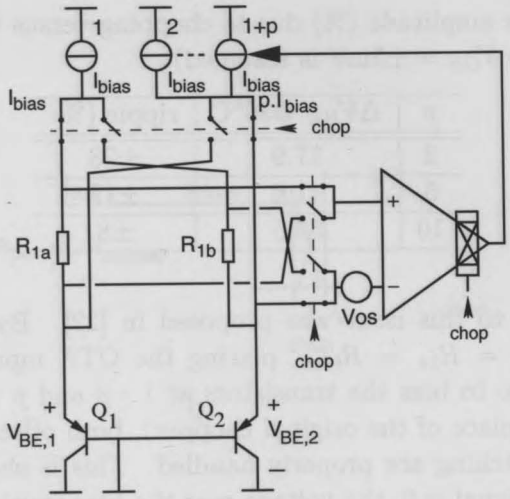


Figure 3.12: Chopping scheme used to compensate the OTA offset.

that actually has to be analyzed. The first term of (3.47) reveals that no matching requirement between Q_L and Q_R nor between R_{1a} and R_{1b} is required, while the second term, also written as $\frac{1}{2}V_T\left(\frac{V_{OS}}{V_T \cdot \ln(p)}\right)^2$, gives the error in V_{BE} due to the OTA's offset.

Table 3.4 summarizes the estimated error in V_{BE} (in °C) at room temperature[7] for three different current ratios p (assuming an OTA's offset $V_{OS} = 5\text{mV}$). Although a current ratio $p = 10$ gives the smallest error, it requires one PNP transistor to be biased with a current 10 times larger than the other. As mentioned, this eases β -mismatch since the current density flowing through the PNP transistors is very different. Additionally, a bias current one order of magnitude higher is definitely not appropriate in terms of power consumption. On the other hand, $p = 2$ sets a circuit with minimum power consumption (given the same unit current) but at the price of the largest error (greater than the whole error budget available). A current ratio of 5 is a better trade-off, but attention has to be paid to the OTA initial offset V_{OS}. Interestingly, the current ratio p in the pre-bias set equal to the current ratio m in the core may help improving β-matching levels by operating pre-bias and core circuits under more similar conditions, i.e. under the same current ratios.

[7]Although the worst case scenario is at $-55°\text{C}$, the best trade-off analysis is not affected.

Table 3.4: Residual error due to OTA's offset evaluated at room temperature for different current ratios p.

p	error in V_{BE} (°C)
2	≈ 0.25
5	≈ 0.05
10	≈ 0.02

3.6.4 Compensation for the finite forward current gain β

As mentioned before, the PNPs are biased through their emitters. Because V_{BE} depends on the collector current instead of on the emitter current, the voltage V_{BE} ends up depending on the current gain (refer to Subsection 2.4.4 in Chapter 2). Since β suffers from process spread, so will V_{BE}, causing sensors to deviate from the average characteristic. Obviously, for high values of β, the collector current approaches the emitter current, i.e. $I_C \approx I_E$, which would minimize the problem. However, since the substrate PNP is a compatible transistor in standard CMOS technologies, its β is relatively small when compared with, for instance, a similar transistor in bipolar or Bi-CMOS technologies.

In order to eliminate or at least to compensate the V_{BE} dependency on β, some circuits have been reported in the literature [45][46]. E.g. in the last, a correcting current component of an auxiliary PNP $I_B = I_E - I_C$ is added to the emitter current of the primary transistor so to help to generate the needed biasing. Because, ideally, the compensation should be implemented without any extra biasing current, so to avoid increasing the front-end supply current, this technique was not considered for implementation in this work. On the other hand, the topology shown in Fig. 3.13 and proposed in [22] satisfies this requirement, since it is implemented simply by adding a resistor R_2 in series with the base of transistor Q_2 without any extra-biasing. The key property behind this method is to generate a bias current that depends on β in such way that, when used in the core, cancels the β-dependency of V_{BE}. Because of the feedback,

$$-V_{BE,1} - V_{R1} + V_{BE,2} + V_{R2} = 0 \tag{3.48}$$

or alternatively

$$\Delta V_{BE,bias} = V_{BE,2} - V_{BE,1} = I_{bias}R_1 - \frac{pI_{bias}}{\beta+1}R_2. \tag{3.49}$$

If the resistor R_2 is set to p times smaller than R_1, i.e. $R_2 = \frac{R_1}{p}$, $\Delta V_{BE,bias}$

results in

$$\Delta V_{BE,bias} = I_{bias} R_1 \left(1 - \frac{1}{\beta+1}\right) = I_{bias} R_1 \left(\frac{\beta}{\beta+1}\right), \tag{3.50}$$

or alternatively, if solved for I_{bias}, in

$$I_{bias} = \frac{\Delta V_{BE,bias}}{R_1 \left(\frac{\beta}{\beta+1}\right)} = \left(\frac{\beta+1}{\beta}\right) \frac{\Delta V_{BE,bias}}{R_1}. \tag{3.51}$$

This current, when applied to bias the PNP transistor in the core results in

$$\begin{aligned} V_{BE} &= V_T \ln\left(\frac{\beta}{\beta+1}\frac{I_{bias}}{I_s}\right) \\ &= V_T \ln\left(\frac{\Delta V_{BE,bias}}{R_1 I_s}\right) \end{aligned} \tag{3.52}$$

which is not dependent on β any more, if a perfect β-match is assumed between the PNP transistors [22]. This means that a $\Delta V_{BE} \approx 41.6$mV (assuming a current ratio $p = 5$ and room temperature) is corrected by a factor $\frac{\beta}{\beta+1}$, resulting in a $\Delta V_{BE,bias} \approx 43.5$mV (for a nominal $\beta = 22$ [28]). This enhanced value for ΔV_{BE} causes I_{bias} to be larger, so to bias the primary PNP with a proper collector current I_C, despite of such terminal being not accessible. In the pre-bias circuit, a nominal (and β-compensated) bias current of $I_{bias} = 125$nA requires bias resistor $R_1 = 348$kΩ. The resistor R_2 is then 69.6kΩ.

The full β-compensation and offset cancellation scheme is shown in Fig. 3.14 (also in Fig. 3.15(a)). It produces a highly accurate PTAT current, which is mirrored to the front-end core, shown in Fig. 3.15(b). The power dissipated is roughly set by the $(1 + p)$ current sources and OTA supply current. No extra current sources specifically for this technique are needed.

3.6.5 Error due to mismatch in the current sources

The bias current I_{bias} generated in the pre-bias circuit is mirrored to the core. While the absolute value of the bias current is not important when generating the voltage ΔV_{BE}, it is for V_{BE} generation. In addition to the errors addressed in former sections, errors in the bias current due to the mismatch of current sources in the core may also contribute to the total spread of the voltage V_{BE}. This is analyzed as follows.

A set of m current sources is used to generate a coarse bias current ranging form 1 to m times I_{bias}, since one current source is left for fine trimming (as

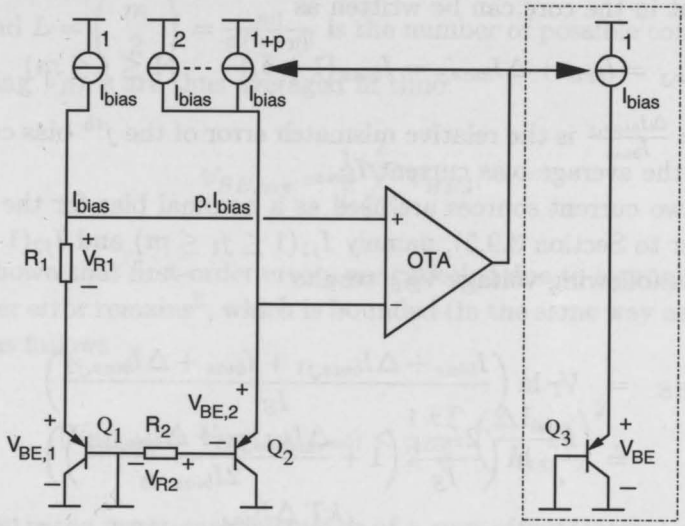


Figure 3.13: β-compensation scheme. The bias current (in the pre-bias) is generated depending on β in such way to cancel the V_{BE} β-dependency of Q_3 (representing the front-end core).

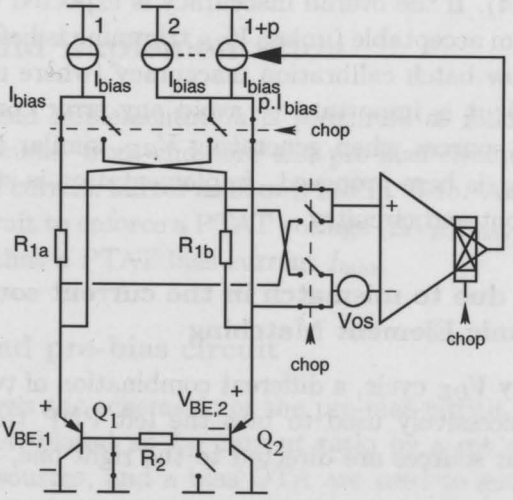


Figure 3.14: Full β-compensation and offset cancellation scheme.

discussed later in Section 3.9). Because the current sources are not ideal, a bias current in the core can be written as

$$I_{bias,j} = I_{bias} + \Delta I_{bias,j} = I_{bias}\,(1 + \delta_j) \qquad (1 \le j \le m) \tag{3.53}$$

where $\delta_j = \frac{\Delta I_{bias,j}}{I_{bias}}$ is the relative mismatch error of the j^{th} bias current with respect to the average bias current I_{bias}.

Since two current sources are used as a nominal bias for the PNP transistor (refer to Section 3.9.5), namely $I_{j1}(1 \le j_1 \le m)$ and $I_{j2}(1 \le j_2 \le m)$, $j_1 \ne j_2$, the following voltage V_{BE} results

$$V_{BE} = V_T \ln\left(\frac{I_{bias} + \Delta I_{bias,j1} + I_{bias} + \Delta I_{bias,j2}}{I_S}\right) \tag{3.54}$$

$$= \frac{kT}{q} \ln\left(\frac{2I_{bias}}{I_S}\left(1 + \frac{\Delta I_{bias,j1} + \Delta I_{bias,j2}}{2I_{bias}}\right)\right) \tag{3.55}$$

$$\approx V_{BE}|_{\Delta I_{bias,j1,2}=0} + \frac{kT}{q}\frac{\Delta I_{bias}}{I_{bias}} \qquad (\Delta I_{bias} \ll I_{bias}) \tag{3.56}$$

where $\Delta I_{bias} = \frac{\Delta I_{bias,j_1} + \Delta I_{bias,j2}}{2}$.

The second term in (3.56) is an approximation to the absolute error introduced in V_{BE} due to an error ΔI_{bias} in the bias current[8] due to a mismatch in the core. If, for example, a current mismatch $\frac{\Delta I_{bias}}{I_{bias}} = 1\%$ is assumed, an absolute error in V_{BE} of about 0.25mV is estimated which is about 0.08°C according to (3.14). If the overall inaccuracy is expected to be below 0.1°C, this error is seldom acceptable (unless V_{BE} trimming is beforehand assumed). However, for a low batch calibration inaccuracy (where no individual trimming is planned), it is important to avoid any error contribution to V_{BE}. DEM of current sources when generating V_{BE}, similar to that used when generating ΔV_{BE}, is here proposed. Implementation is straightforward and uses the same front-end circuitry.

3.6.5.1 Error due to mismatch in the current sources with Dynamic Element Matching

With DEM, every V_{BE} cycle, a different combination of two current sources, among m, is successively used to bias the left PNP transistor, while the remaining current sources are directed to the right one, resulting in a V_{BE} given by

$$V_{BE,l} = V_T \ln\left(\frac{\sum_{L=l} I_{bias,l}}{I_S}\right) \qquad (1 \le l \le L) \tag{3.57}$$

[8] In fact an averaged error given the errors of two current sources.

where $\sum_{L=l} I_{bias,l}$ is the sum of two current sources relatively to the l^{th} combination and $L = \begin{pmatrix} m \\ 2 \end{pmatrix} = \frac{m!}{p!(m-p)!}$ is the number of possible combinations. The resulting V_{BE}'s are thus averaged in time:

$$V_{BE,avg} = \frac{1}{L} \sum_{l=1}^{L} V_{BE,l}. \tag{3.58}$$

It can be shown that first-order errors are canceled due to averaging, while a second-order error remains[9], which is bounded (in the same way as developed in (3.22)) as follows

$$|V_{BE,avg} - V_{BE}|_{\Delta I_{bias}=0}| < \frac{1}{2} \frac{kT}{q} \left(\frac{\Delta I_{bias}}{I_{bias}} \right)^2, \tag{3.59}$$

where $\frac{\Delta I_{bias}}{I_{bias}}$ is the worst-case mismatch of a sum of two current sources.

This shows that now, for the same assumed relative mismatch of 1%, the error contribution due to current sources mismatch can be neglected (i.e. error below mK). Obviously, the spread in the saturation current I_S dominates, such that the effect of DEM when generating V_{BE} is not easily observed. This technique is effective though, while not visible due to another major cause of V_{BE} spread.

3.7 Front-end implementation

The sensor front-end implementation is presented as follows. It is mainly built from two circuits: front-end core and pre-bias circuit, linked together by means a shared current mirror as shown in Fig. 3.15. An OTA is required in the pre-bias circuit to enforce a PTAT voltage ($\Delta V_{BE,bias}$) over the resistor $R_{1a,b}$, generating thus a PTAT bias current I_{bias}.

3.7.1 Front-end pre-bias circuit

Figure 3.15(a) shows the schematic of the pre-bias circuit. Two PNP transistors with area A, biased at $1{:}p$ current ratio by a set of $(1+p)$ PMOS cascoded current sources, and a bias OTA are used to generate $\Delta V_{BE,bias}$. This voltage is applied over a resistor $R_{1a,b}$ to generate a PTAT bias current I_{bias}. A *chop* signal is used for OTA chopping as explained.

[9] Third and higher-order errors were neglected.

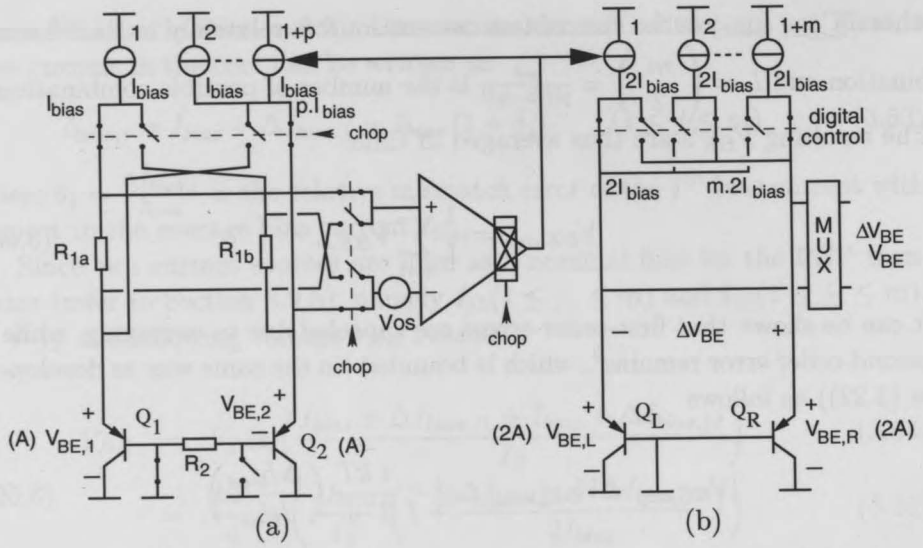


Figure 3.15: Front-end schematic. (a) The pre-bias circuit and (b) the front-end core.

3.7.2 Front-end core

Figure 3.15(b) shows the schematic of the front-end core. Two PNP transistors with area $2A$, biased at $1 : m$ current ratio by a set of $(1+m)$ PMOS cascoded current sources ($2I_{bias}$) are used to generate ΔV_{BE} and V_{BE} (selected via a multiplexer). While in theory only one PNP transistor is necessary [40], the generation of ΔV_{BE} would require in this case two cycles. The lack of any PNP transistor mismatch is thus counter balanced by a more complex circuitry, a longer ΔV_{BE} cycle and sampling errors related to V_{BE} storing. A digital control is used for chopping, DEM and trimming as discussed.

3.7.3 The bias OTA

An OTA is required to enforce a $\Delta V_{BE,bias}$ over the bias resistor $R_{1a,b}$, so a PTAT current as given in (3.29) is generated. Since the OTA offset is a major source of error, it is chopped.

Fundamental requirements

Bandwidth, DC gain (A_{DC}) and power consumption are the most important parameters of the bias OTA. Proper bandwidth is required, so the amplifier can settle to the required accuracy within a given time for the specified capacitance load C_L (here given mainly by the estimated total gate-source C_{GS}

Table 3.5: Bias OTA main target specifications. The load capacitance C_L was estimated from simulations.

Parameter	Value
DC gain (dB) over corners	≥ 65
GBW= $\frac{g_m}{2\pi C_L}$ @ $C_L \approx 10$pF (kHz)	≥ 100
phase margin (°)	≥ 60
voltage supply (V)	$2.5 - 5.5$
total supply current (μA)	≤ 2.0

capacitance of all core and pre-bias current-sources transistors connected to the OTA output). Concerning the DC gain, a value high enough is necessary to guarantee accuracy, so for instance, errors in V_{BE} as given by (3.39) are negligible. Both bandwidth and DC gain are dependent on the amplifier biasing level. A design decision of having an OTA with a total supply current of 2μA was established to allow some margin to have a front-end total supply current not greater than 5μA. This value is much smaller than that in prior art and most of related works. Table 3.5 summarizes the main target specifications of the OTA.

Topology

A folded cascode OTA with a chopped output is proposed. Since cascoding increases the amplifier output impedance, the required DC gain is now achievable without increasing the g_m of the input pair transistor through a higher bias current. Furthermore, because of the folded-cascode current-follower stage, the amplifier provides an output-voltage swing independent of the input common mode [47], so an appropriate bias level is set to the core and pre-bias current sources regardless of the input common mode set by the PNP transistors.

However, not only the gain A_{DC} but the gain-bandwidth product GBW depends on the tail current. In the later case, the only option is to minimize the output capacitance load seen by the OTA, so to have the required GBW without increasing the tail current of the amplifier.

Because of the OTA's offset V_{OS}, the chopping technique [43] is used. While the input chopper is implemented externally, the output chopper is built internally, as shown in Fig. 3.16. The initial offset voltage V_{OS}, nonetheless, is minimized by designing the OTA input pair deeply in weak inversion, so matching is improved and the amplifier's offset reduced [41][11]. The weak inversion also helps to achieve the required g_m with a smaller bias.

Ripple due to chopping

While chopping is used to cancel the amplifier offset, it causes ripple in the bias current. A larger current ratio p can help reducing the ripple and the error associate with it. But a larger p means more current sources in the pre-bias circuitry which increases the front-end supply current, as discussed.

Ripple filtering techniques might be considered to limit the ripple effect after chopping. The conventional low pass filtering, however, using a large capacitor at the output of the OTA (combined with its limited bandwidth) is not indicated. First, because is requires a prohibitively large capacitor, difficult to be integrated, and second, because it degrades the current ratio: the mirror is not capable of defining accurately the current ratio while the currents are varying (due to different loads the circuit sees in each side). More recent approaches reduce the ripple due to chopping by using sample-and-hold filters [48][49] which give better results. They might be considered in the case where the error due to ripple, given a small p, is still too high after chopping.

Design and implementation

The design of the bias OTA follows the well-known and classic OPAMP design. Given the specified total supply current for the OTA, $4I_B$ (refer to Fig. 3.16) is set to 1.5μA while the remaining 0.5μA is used for biasing (which gives a total 2.0μA amplifier). The input pair transistors are then biased with $I_B = 0.375\mu$A in weak inversion ($\frac{g_m}{I_D} \approx 20\text{V}^{-1}$, i.e $g_m \approx 8\mu\text{S}$). For an output load capacitance $C_L = 10\text{pF}$ (estimated from simulations with the actual current mirror), a GBW of 100kHz is achieved.

Because of the OTA's low tail current, thermal noise was estimated. According to [50], the power spectral density (PSD) of a simple OPAMP (valid for the folded cascode if the noise due to the input transistors is dominant) can be approximated by

$$V_n^2(f) = \frac{16}{3}\frac{kT}{g_m}\left[\frac{\text{V}^2}{\text{Hz}}\right] \tag{3.60}$$

where k is the Boltzmann constant and g_m is the transconductance of the input pair transistors. The total noise contribution is evaluated by taking this PSD shaped by the OTA first-order low-pass ($\frac{1}{2\pi}\frac{g_m}{C_L}$) filter resulting in

$$V_n^2 = \frac{4}{3}\frac{kT}{C_L}. \tag{3.61}$$

At $T = 400\text{K}$, for $C_L = 10\text{pF}$, the estimated total noise contribution is about $30\mu\text{V}_{rms}$, or roughly 0.01°C.

Simulation Results

Figure 3.17 shows the simulated frequency response of the bias OTA for a load capacitance $C_L = 10\text{pF}$ and a tail current of $0.75\mu\text{A}$. The nominal amplifier's DC gain is 85dB, while the nominal gain-bandwidth product (GBW) is 100kHz, with a phase margin higher than 60°. The OTA draws a total current of $2.0\mu\text{A}$ (about 40% of the total front-end current) as specified.

3.7.4 Front-end total supply current summary

The front-end total supply current can be determined after the current ratios m and p, core and pre-bias currents and bias OTA current were set based on specification requirements. For an inaccuracy smaller than 0.1°C, operating range up to 130°C, and actual PNP transistor sizes, the total supply current found is below $5\mu\text{A}$. Obviously, this value is temperature dependent (determined here for 30°C) since currents in this circuit are mostly PTAT. Although improvements can be always expected, this value represents to the best knowledge of the author a very good approximation of the minimum current required given the sensor specification and the addressed architecture. Table 3.6 summarizes the supply current of all front-end building blocks. Assuming $V_{DD} = 2.5\text{V}$, the estimated total power consumption (at 30°C) is $12\mu\text{W}$. Such power dissipation allows labeling of this front-end as a low-power circuit, and suggests its use to build very-low power sensors.

3.8 Current mirror

The bias current accurately generated in the pre-bias circuit has to be mirrored to the core. Because such function was found critical in the front-end as already shown, a more detailed analysis is carried out here. The mirroring procedure, because it is implemented with non-ideal current sources, often introduces errors. Cascoding as previously mentioned diminishes the effect due to the current source finite output impedance. However, mismatches of parameters such μ, C_{ox}, W, L and V_{TH} result in mismatches between the drain currents (for a given V_{GS}) of two or more nominally-identical current sources [41], e.g. from the pre-bias to the core. Additionally, because of the PTAT bias current (which implies a temperature-dependent bias point), matching levels (and errors) are also temperature-dependent. This situation is addressed as follows starting with typical mismatching models. A trade-off among the three main performance specifications, i.e. power consumption,

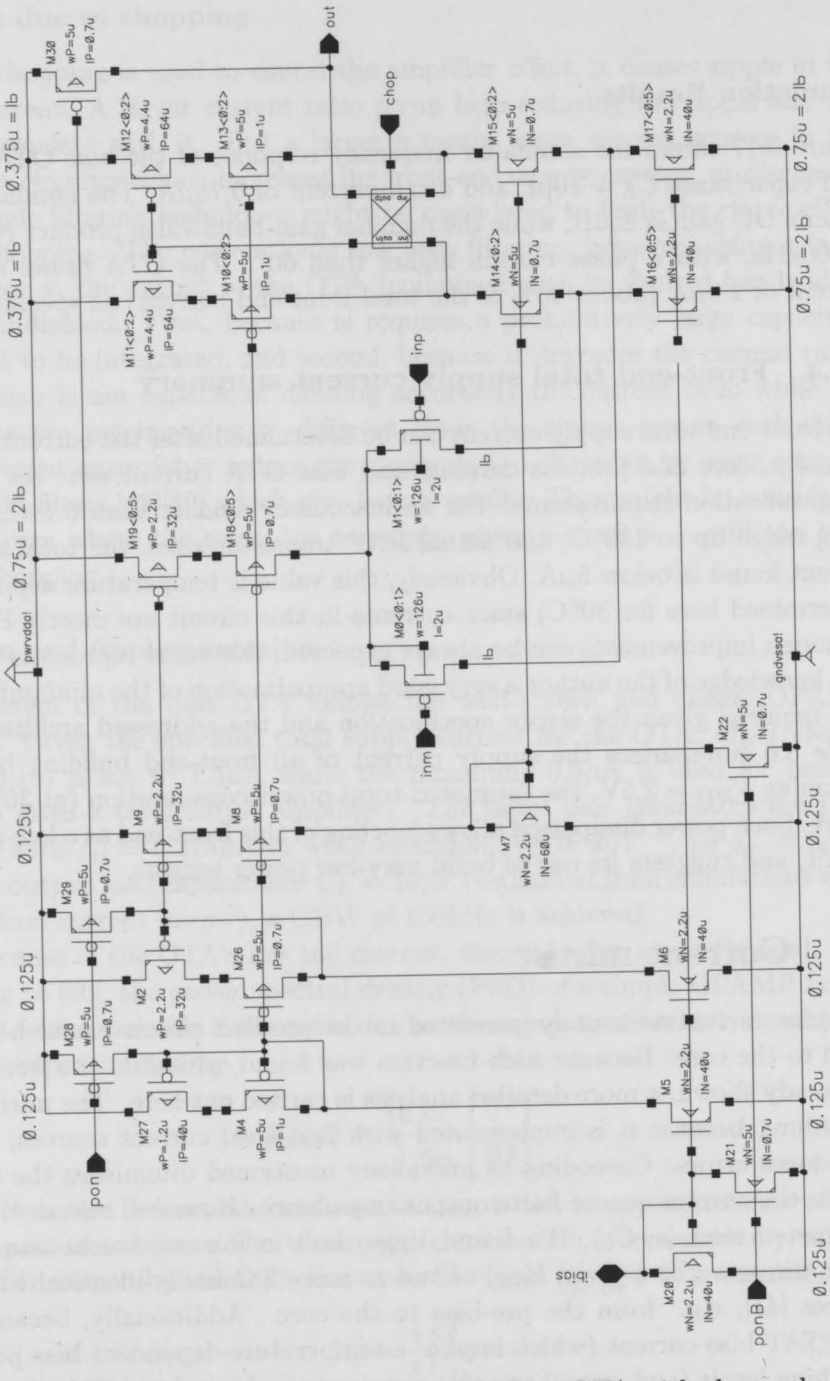


Figure 3.16: Schematic of the folded-cascode bias OTA with chopped output current mirror.

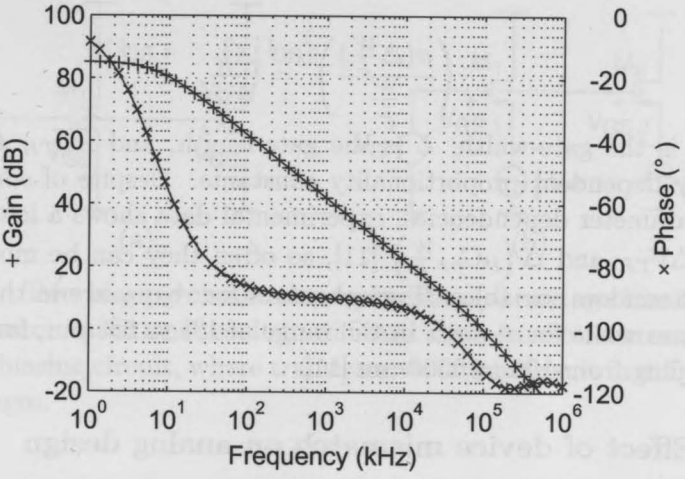


Figure 3.17: Simulated frequency response of the bias OTA: DC gain and phase.

Table 3.6: Front-end total supply current at 30°C .

front-end circuit	current
core (1:5)	$6 \times (2 \times 125\text{nA})$
pre-bias (1:5)	$6 \times 125\text{nA}$
bias OTA	$2\mu\text{A}$
other bias	$0.5\mu\text{A}$
TOTAL	$4.75\mu\text{A}$

accuracy (or error) and speed (conversion time or bandwidth BW) is further developed.

3.8.1 Device mismatch models

The mismatch of two closely spaced, identically designed MOS transistors has been extensively investigated. The experimental data shows that threshold voltage differences ΔV_{TH} and current factor differences $\Delta\left(\mu C_{ox}\frac{W}{L}\right)$ are the dominant sources underlying either the drain-source current or the gate-source voltage mismatch for a matched pair of MOS transistors [11]. These random differences have a normal distribution with zero mean and a variance dependent on the device area WL [51]:

$$\sigma^2\left(\Delta V_{TH}\right) = \frac{A_{VTH}^2}{WL} \tag{3.62}$$

$$\left(\frac{\sigma(\Delta\frac{W}{L})}{\frac{W}{L}}\right)^2 = \frac{A_K^2}{WL} \tag{3.63}$$

where W is the gate-width, L is the gate-length, and A_{VTH} and A_K are technology dependent proportionality constants. Despite of some common process parameter dependencies, experimental data shows a low correlation between ΔV_{TH} and $\Delta\left(\mu C_{ox}\frac{W}{L}\right)$ [11], so often they can be modeled as independent random variables. Typical values for A_{VTH} are in the range of 2 to 40 mVμm while for A_K are in the range of 1% to 5% μm, for technology nodes varying from 100 to 2500 nm [11].

3.8.2 Effect of device mismatch on analog design

Two types of errors are typically of interest in analog design. For the voltage biased pair in Fig. 3.18(a), the current matching error $\Delta I_{DS} = I_{DS,2} - I_{DS,1}$, between the drain-source currents $I_{DS,1}$ and $I_{DS,2}$ of M_1 and M_2 respectively, determines the circuit accuracy. The gate-source voltage error $\Delta V_{GS} = V_{GS,2} - V_{GS,1}$ sets the accuracy in a current biased configuration, as shown in Fig. 3.18(b). These errors can be expressed in terms of ΔV_{TH} and $\Delta\left(\mu C_{ox}\frac{W}{L}\right)$ of the matched device pair in saturation [41][11][10]

$$\left(\frac{\sigma(\Delta I_{DS})}{I_{DS}}\right)^2 = \left(\frac{\sigma(\Delta\frac{W}{L})}{\frac{W}{L}}\right)^2 + \left(\frac{g_m}{I_{DS}}\right)^2 \sigma^2(\Delta V_{TH}) \tag{3.64}$$

$$\sigma^2(\Delta V_{GS}) = \frac{1}{\left(\frac{g_m}{I_{DS}}\right)^2}\left(\frac{\sigma(\Delta\frac{W}{L})}{\frac{W}{L}}\right)^2 + \sigma^2(\Delta V_{TH}) \tag{3.65}$$

where $\frac{g_m}{I_{DS}}$ depends on the bias point (sometimes used in place of $(V_{GS} - V_{TH})$).

The best current matching is achieved by using a bias point with a small $\frac{g_m}{I_{DS}}$ or a large gate-source overdrive, i.e. deep into strong inversion while the best voltage matching is achieved for a bias point with a large $\frac{g_m}{I_{DS}}$ or a low gate-source overdrive, i.e. operation towards weak inversion. In both cases, regardless of the biasing point, errors are smaller for smaller parameter variations, which is achieved by making the devices larger (with greater area) (refer to (3.62) and (3.63)).

It is important to note that for gate-source overdrive voltages of some hundreds of mV, i.e. $V_{GS} - V_{TH} < 0.5$V, the contribution of the threshold

[10] Here only the current error in a voltage biased circuit is of interest. Nonetheless, for sake of completeness, however, both ΔI_{DS} and ΔV_{GS} errors are presented.

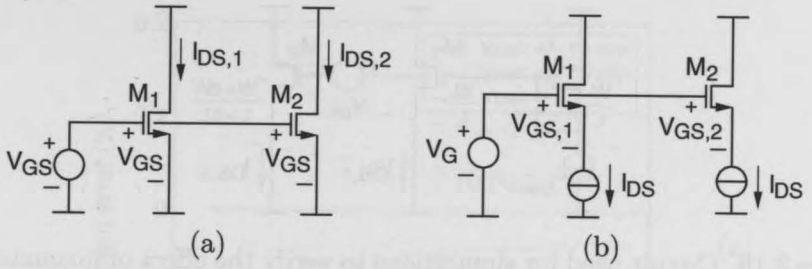


Figure 3.18: Two biasing circuit arrangements: (a) voltage biasing circuit, where transistor variations result in different drain-source currents I_{DS} and (b) current biasing circuit, where transistor variations result in different gate-source voltages.

voltage mismatch ΔV_{TH} to the voltage or current errors is dominant over the current factor mismatch $\frac{\Delta\left(\frac{W}{L}\right)}{\frac{W}{L}}$ contribution [11][51].

Furthermore, in the particular case of a voltage biasing circuit, the dominant term ΔV_{TH} in the current error as given in (3.64) appears multiplied by $\frac{g_m}{I_{DS}}$ which depends on the temperature since the bias current is PTAT[11]. Therefore, regardless of the factor $\frac{\Delta\left(\frac{W}{L}\right)}{\frac{W}{L}}$, the term $\left(\frac{g_m}{I_{DS}}\right)\Delta V_{TH}$ in (3.64) is then of main concern when determining the errors due to mismatch in current mirrors over temperature.

3.8.3 Current mirroring over temperature

In order to verify the mirroring characteristic over temperature, simulations over temperature were performed. The circuit schematic used for this purpose is shown in Fig. 3.19. An ideal PTAT current source with a temperature dependency close to that used in the front-end was used. Simulation results of the relative current error due to ΔV_{TH} (introduced by means of an ideal voltage source V_{OS}) and $\Delta\left(\mu C_{ox}\frac{W}{L}\right)$ (introduced by means of variation in W and L) are shown in Fig. 3.20(a), (b) and (c).

Fig. 3.20(a) shows that the current error due to the mismatch in the threshold voltage of two transistors M_1 and M_2 is strongly temperature dependent. This is predicted by (3.64). Because the current is PTAT, which is achieved by modifying the bias point (i.e. the value of $\frac{g_m}{I_{DS}}$), higher mismatch is observed for lower temperatures where the value of $\frac{g_m}{I_{DS}}$ is relatively larger.

[11]The transconductance (in strong inversion) $g_m = \sqrt{2\mu(T)C_{OX}\frac{W}{L}\cdot I_{DS}(T)}$ is essentially constant with the temperature, since the mobility decreases with temperature (i.e. $\mu(T) \approx \mu_o T^{-n}$ where $n \approx -\frac{3}{2}$ [41]), partially compensating the PTAT bias current I_{DS}.

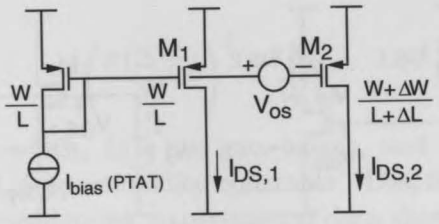


Figure 3.19: Circuit used for simulations to verify the effect of mismatch over temperature in a current mirror.

For higher temperatures, when the bias current gets larger, $\frac{g_m}{I_{DS}}$ gets smaller, consequently a lower mismatch is obtained.

The errors due to variation in the transistor width or length on the other hand are not temperature dependent. This is shown in Fig. 3.20(b) and (c), also in good agreement with (3.64). The small temperature dependence observed for variations in ΔW is indeed due to ΔV_{TH} caused by ΔW because of the cross-correlation between V_{TH} and W [41]. Additionally, errors due to variations in ΔW and ΔL (a 0.1% variation is assumed) are about 10 times smaller than the errors due to variations in ΔV_{TH} (offset voltages up to $\pm$10mV were assumed) for transistors of nominal sizes of $\frac{W}{L} = \frac{10\mu m}{40\mu m}$.

3.8.4 Impact of device mismatch on the performance trade-offs of front-end

Three main performance specifications in most of analog circuits are often the operation speed (or bandwidth BW) of the circuit, its accuracy (or its inverse, the error or inaccuracy) and the associated power consumption [52]. This is not different for the sensor front-end. Since the primary goal of this work is to reduce the power consumption to minimum values set by accuracy and speed requirements, it is important to find a relation among them so appropriate trade-offs can be established.

The BW of a circuit can be related to the transit frequency f_T of the MOS devices, i.e. the BW is proportional to $f_T = g_m/(2\pi C_{GS})$ where the exact gate-source capacitance C_{GS} depends on the particular circuit: it is $(C_{GS1} + C_{GS2})$ in the case of a 1 : 1 current mirror or the sum of all capacitances C_{GS} at the output of the bias OTA (appearing as an equivalent load capacitance C_L). Regardless of the case, the gate-source capacitance C_{GS} is always written as a function of the oxide capacitance C_{ox} and device area, i.e. $C_{GS} \propto C_{ox}WL$.

Because the power consumption P of a circuit operating from a power

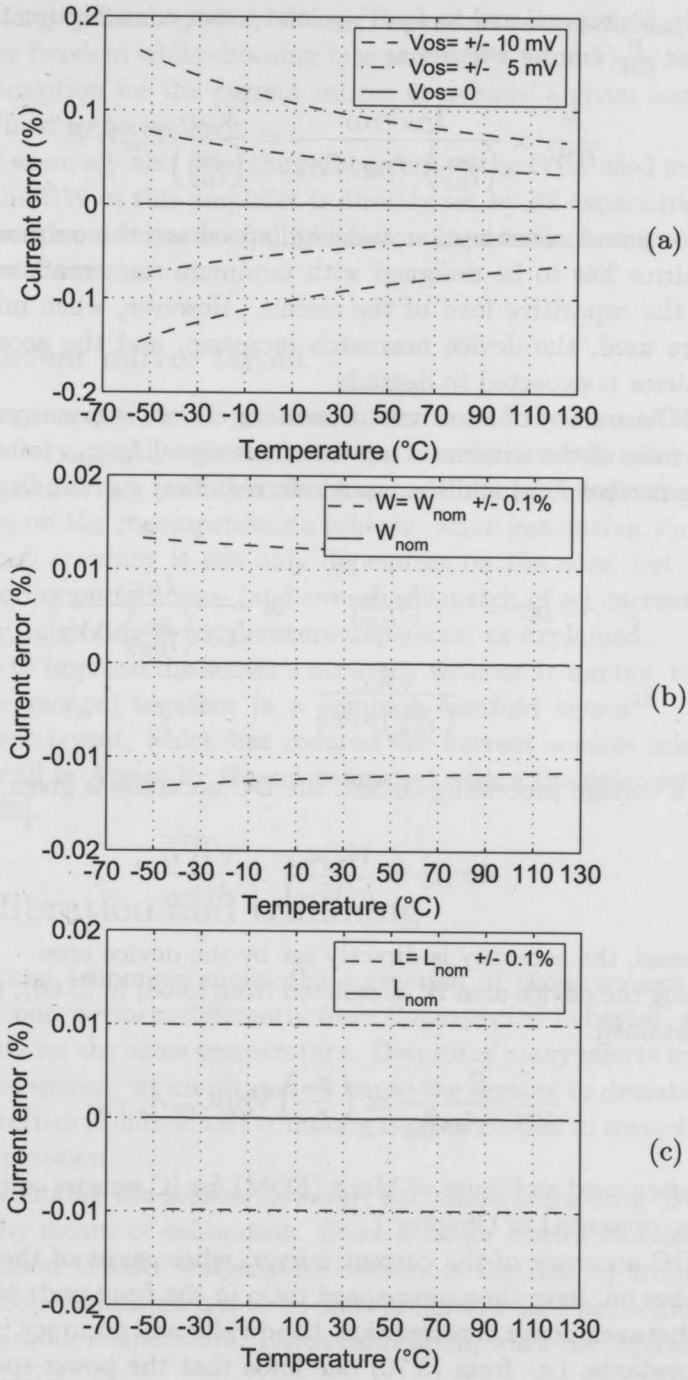


Figure 3.20: Current error (%) calculated as $(I_{D2} - I_{D1})/I_{D1}$ as a result of mismatch in (a) V_{TH} (b) W and (c) L.

supply V_{DD} is proportional to $I_{DS}V_{DD}$, the power consumption to the bandwidth ratio $\frac{P}{BW}$ can be written as

$$\frac{P}{BW} \propto \frac{I_{DS}V_{DD}}{\left(\frac{g_m}{I_{DS}}\right).I_{DS}/C_{GS}} = \frac{V_{DD}}{\left(\frac{g_m}{I_{DS}}\right)} C_{ox}WL. \tag{3.66}$$

If a low power consumption and high speed are the only concerns, the current mirror has to be designed with minimum size transistors, so as to minimize the capacitive load of the circuit. However, when minimum size devices are used, the device mismatch increases, and the accuracy of the current mirror is expected to degrade.

The DC accuracy of a current processing circuit, e.g. current mirror, is set as the ratio of the maximum input current signal $I_{in,rms}$ (often a fraction of the bias current I_{DS}) and the input referred offset current I_{OS} as given by [11]

$$Acc_{DC} = \frac{I_{in,rms}}{\sigma(I_{OS})} \propto \frac{I_{DS}}{I_{DS}\left(\frac{g_m}{I_{DS}}\right)\Delta V_T} \tag{3.67}$$

$$\propto \frac{\sqrt{WL}}{\left(\frac{g_m}{I_{DS}}\right)A_{VTH}} \tag{3.68}$$

while for a voltage processing circuit, the DC accuracy is given by

$$Acc_{DC} = \frac{V_{in,rms}}{\sigma(V_{OS})} \propto \frac{\sqrt{WL}}{A_{VTH}}. \tag{3.69}$$

In both cases, the accuracy is directly set by the device area.

By using the device area $W \cdot L$ isolated from (3.68) in (3.66), the following ratio is obtained:

$$\frac{P}{BW \cdot Acc_{DC}^2} \propto \left(\frac{g_m}{I_D}\right) V_{DD}C_{ox}A_{VTH}^2. \tag{3.70}$$

which if often used as Figure-of-Merit (FOM) for IC sensors comparison[12] as the survey presented in Chapter 1.

The DC accuracy of the current mirror, where most of the sensor's accuracy relies on, fixes the power-speed ratio in the front-end. Moreover, the trade-off between power consumption, bandwidth and accuracy is set by technology constants, i.e. from (3.70) one finds that the power-speed-accuracy

[12]Sometimes written as $P \cdot T_C \cdot InAcc^2$,where T_C is conversion time and $InAcc^2$ the square of inaccuracy or relative error.

trade-off is largely fixed by a technology dependent constant $C_{ox}A_{VTH}^2$. Despite of some freedom while choosing bias point ($\frac{g_m}{I_{DS}}$) and V_{DD}, the required power consumption for the current mirror to achieve a given accuracy and speed is defined by technology.

The DC accuracy also fixes the tail current of the OTA used in the front-end, since the BW of this amplifier is directly set by its capacitive load, the total gate-source capacitance of all transistors used to implement the current mirror under discussion.

3.8.5 Current mirror layout

The current generated in the pre-bias is mirrored to the core. Although the sensor's accuracy greatly relies only on the bias current ratio m in the core, when generating ΔV_{BE} (and this can be successfully achieved using DEM), it depends also on the β-compensation scheme, when generating V_{BE}. In this case, however, accuracy is not only dependent on the core, but relies also on the mirroring performance (and overall mismatch of all current sources). Additionally, mirroring is temperature dependent as explained.

In order to improve the sensor's accuracy without trimming, the current sources were merged together in a common centroid layout[13]. This new current mirror layout, which has reduced the current sources mismatch, is shown in detail in Appendix H, and compared with a the prior-art layout[14], as used in [22].

3.9 Calibration and trimming

Calibration and trimming allows the correction of those sensors that, for different reasons, perform differently from the expected behavior, giving different outputs for the same temperature. Despite of many efforts to diminish the sources of spread, which ultimately cause the sensors to deviate from the ideal characteristic, (individual) trimming is often needed to correct or adjust the sensor operation.

Before trimming the sensor, however, its temperature error needs to be established by means of calibration. Since a major contributor to the total cost of precision CMOS temperature sensors is the cost of trimming and calibration [53], calibration has to be done preferable once per sensor sample, at or close to room temperature. Batch calibration, when the typical behavior

[13] Additionally, inter-DEM (DEM of all current sources, and not only of those of the core – intra-DEM) might be considered.

[14] This layout was kindly supplied by the author.

of all samples (population) is estimated by means of the calibration of some samples can provide significant cost savings but this is usually at the expense of an equally significant loss of accuracy.

Trimming requires extra circuitry, not to mention the fact that it has to be effective over the entire temperature range instead of only at the calibration point. This requires the understanding of the trimming mechanism, and parameters involved, so the sensor behavior after trimming is reasonably known. For instance, to trim a linear (temperature) characteristic, two points have to be fixed (which requires calibration at two different points). Fortunately, in the case of bandgap references, one point is already fixed by physical constants: at 0K the extrapolated curve of V_{BE} (and V_{REF}) intersect the vertical axis at the value of V_{BE0} (the linearly extrapolated value of V_{BE} at $T = 0$K) [32]. For this reason, the V_{BE} characteristic can only rotate around V_{BE0}, meaning that a single trimming is enough to adjust this curve, modifying the sensor output. It is worth mentioning that the success of the one-step trimming is lost if any other non PTAT error sources are present.

Temperature calibration, performed at room temperature (30°C), using calibrated Pt100 resistors as absolute temperature references, instead of for instance voltage calibration [33] was chosen for this work. The slightly superior results of such calibration technique justified this choice.

Following the same understanding as in [22], and agreeing with the theory presented in Chapter 2, the errors in V_{BE} due to processing spread (and fabrication spread, e.g. due to mechanical stress) were taken as dominant (if errors in the readout circuitry are assumed negligible). Therefore V_{BE} trimming was chosen as the trimming mechanism in this work.

3.9.1 V_{BE} trimming by PNP bias-current adjusting

Trimming in general adjusts the operation of a sensor by modifying a given circuit parameter, e.g. a transistor or resistor bias current or voltage, width or length, etc. There are several methods of doing this [22][32]. One in particular is by trimming the voltage V_{BE}. For instance, the voltage V_{BE} can be trimmed either by modifying the PNP transistor bias current I_{bias} or saturation current I_S (by altering the transistor emitter area). In both cases, the current density flowing through the transistor is effectively modified, which causes V_{BE} to change, allowing proper adjustment.

Trimming V_{BE} by modifying the transistor area is often implemented by adding transistors in parallel with the main transistor (usually following some scaling) [27][32]. The main drawback concerns the switches in series with the parallel-added transistors, whose voltage drop add to the transistors base-emitter voltages.

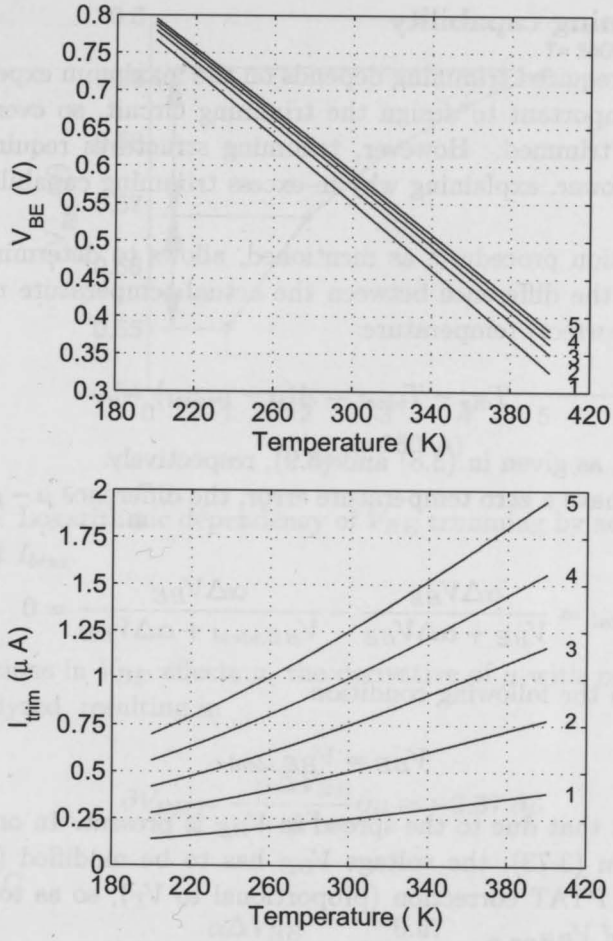


Figure 3.21: (a) Trimming of V_{BE} by adjustment of the bias current I_{bias}(the five coarse modes are shown). (b) The five PTAT bias currents used for coarse trimming, defining the modes 1 to 5. The nominal bias current is set to two unit current, i.e. $I_{bias}(0.5\mu\text{A}, T = 300\text{K})$.

Trimming V_{BE} by modifying the bias current I_{bias} (as shown in Fig. 3.21) on the other hand, does not suffer with the voltage drop across the switches, since the switches are in series with the high-impedance output of the current sources (refer to Fig. 3.23 for circuit details).

3.9.2 Trimming capability

The amount of required trimming depends on the maximum expected sensor spread. It is important to design the trimming circuit, so even the worst sensors can be trimmed. However, trimming structures require area and may consume power, explaining why in-excess trimming capability must be avoided.

The calibration procedure, as mentioned, allows to determine the temperature error, the difference between the actual temperature reading T_{dig} and the ideal (correct) temperature:

$$T_{dig} - T_{ideal} = A(\mu - \mu_{ideal}) \tag{3.71}$$

with μ and T_{dig} as given in (3.8) and (3.9), respectively.

In order to have a zero temperature error, the difference $\mu - \mu_{ideal}$ has to be nulled, i.e.

$$\mu - \mu_{ideal} = \frac{\alpha \Delta V_{BE}}{V_{BE} + \alpha \Delta V_{BE}} - \frac{\alpha \Delta V_{BE}}{V_{BE,ideal} + \alpha \Delta V_{BE}} = 0 \tag{3.72}$$

which results in the following condition

$$V_{BE} = V_{BE,ideal} \tag{3.73}$$

if no errors but that due to the spread in V_{BE} is present. In order to meet the condition in (3.73), the voltage V_{BE} has to be modified (named now $V_{BE.trim}$) by a PTAT correction (proportional to V_T), so as to correct the PTAT spread of V_{BE}

$$V_{BE,trim} = V_{BE} + V_{correction}, \tag{3.74}$$

ideally equal to $V_{BE,ideal}$. In practice, the amount of correction depends on the maximum spread of V_{BE}, which is difficult to estimate. According to [28], a maximum spread of ± 13mV ($|\partial V_{BE}| \approx 1V_T$) is expected for a PNP transistor (vpnp460, $460\mu m^2$). This is about ± 3.25°C at 300K since $\frac{\partial T}{\partial V_{BE}} \approx 0.25$°C/mV. For a smaller PNP, a larger spread might be observable. Fortunately, with a set of $(1+m) = 6$ current sources, trimming up to ± 6°C (about ± 20mV) is possible.

An estimation of the variation of the temperature with the bias current can be obtained. First the derivative of the temperature ($T_{dig} = A\mu + B$) with respect to μ is evaluated, resulting in

$$\partial T = A \partial \mu. \tag{3.75}$$

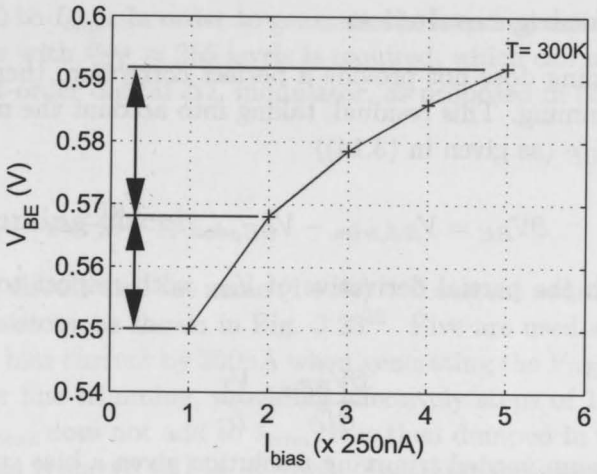


Figure 3.22: Logarithmic dependency of V_{BE} trimming by adjustment of the bias current I_{bias}.

Since variations in V_{BE} affects μ, the derivative of μ with respect to V_{BE} is further analyzed, resulting in

$$\partial V_{BE} = -\frac{\alpha \Delta V_{BE}}{\mu^2} \partial\mu \approx -2.37.\partial\mu \tag{3.76}$$

since at 27°C

$$\mu = \frac{\alpha \Delta V_{BE}}{V_{BE} + \alpha \Delta V_{BE}} \approx \frac{0.67}{1.2} = 0.531. \tag{3.77}$$

The derivative of V_{BE} (as given in (3.1)) with respect to I_{bias} provides the last needed relation to estimate the variation in temperature given a variation in the bias current:

$$\partial V_{BE} = \frac{V_T}{I_{bias}} \partial I_{bias}. \tag{3.78}$$

For example, assume $T = 300\text{K}$ and a $\partial I_{bias} = 250\text{nA}$ (e.g. a jump in the bias current from 500nA to 750nA). Since the geometric average of I_{bias} is 612.37nA, then $\partial V_{BE} = \frac{25.87\text{m}}{612.37\text{n}}.250\text{n} = 10.56\text{mV}$ which means $\partial\mu = -4.45\text{m}$ and $\partial T = -2.58°\text{C}$. Because V_{BE} and I_{bias} are related by a logarithmic function, though, attention must be paid so a nearly-balanced trimming range is available. This is depicted in Fig. 3.22.

3.9.3 Trimming resolution

Because trimming does not provide a perfect correction, there is a residual error after trimming. This residual, taking into account the maximum error tolerated in V_{BE} (as given in (3.14))

$$\partial V_{BE} = V_{BE,trim} - V_{BE,ideal} \leq \frac{V_{REF}}{T}\varepsilon_T \tag{3.79}$$

combined with the partial derivative of V_{BE} with respect to I_{bias} (finite β neglected)

$$\frac{\partial V_{BE}}{\partial I_C} = \frac{V_T}{I_C} \tag{3.80}$$

sets the minimum needed trimming resolution given a bias current I_{bias} and an error contribution ε_T

$$\partial I_C \leq \frac{I_{bias}}{V_T}\frac{V_{REF}}{T}\varepsilon_T. \tag{3.81}$$

If $I_{bias} = 0.5\mu$A, (the *nominal* bias current), $\varepsilon_T = 0.01°$C, a maximum bias variation below 1nA is required (at 300K, $V_{REF} = 1.2$V). This specification is relaxed for a larger error contribution. Even in this case though, such high resolution is difficult to obtain with an integer number of current sources. Bitstream modulation is used to provide the required resolution.

3.9.4 Bitstream trimming

Bitstream trimming [54] allows to define small steps of current, providing conditions for high resolution trimming. The bias current of one current source is switched on and off, generating a current between 0 or I_{bias} depending on the output of a digital modulator. Such modulator produces an output signal whose time average is programmable. The same happens with the generated current I_{fine} in this case given by

$$I_{fine} = \frac{\Sigma t_{on}}{\Sigma t_{on} + \Sigma t_{off}} I_{bias} = \frac{1}{1 + \frac{\Sigma t_{on}}{\Sigma t_{off}}} I_{bias}. \tag{3.82}$$

The total trimmed current is then given by

$$I_{trim} = I_{bias} \times j + I_{fine} \qquad (1 \leq j \leq m = 5) \tag{3.83}$$

where the first term refers to the coarse trimming while the second refers to the bitstream trimming. The coarse trimming generates a current from 1 up to 5 I_{bias} (in steps of I_{bias}) while the bitstream or fine trimming generates a

current from 0 to I_{bias}. In order to generate the required resolution (1nA), a D/A converter with $\frac{I_{bias}}{1\text{nA}} \approx 255$ levels is required, which can be implemented by an 8-b first-order digital $\Delta\Sigma$ modulator, as proposed in [22].

3.9.5 Trimming circuit

The trimming circuit uses the same $(1+m) = 6$ current sources used to bias the PNP transistors, as shown in Fig. 3.23[15]. Five are used as coarse steps, modifying the bias current by 250nA when generating the V_{BE} voltage, while one is used for fine trimming, providing effectively steps of 1nA. When the unit current I_{bias} does not add to I_{trim}, it is then dumped in the other PNP transistor. This is shown in Fig. 3.23(a). The same dumping mechanism is used with the fine trimming (3.23(b)).

As a design choice, the system was set to present a superior performance (where V_{BE} is generated with the smallest non-linearity, refer to next section) using two current sources (coarse mode 2) as nominal bias for the PNP that generates V_{BE}. This allows one coarse step down and three up when trimming. The main reason is to equalize the trimming capability, as shown in Fig. 3.22.

3.9.6 Other trimming methods

Digital trimming methods are very attractive in the sense that they allow the adjustment of the temperature reading $T_{dig} = A\mu + B$ externally[16], usually at several temperatures (which in terms of design might mean more relaxed specifications). Because in general plenty of processing resources are externally available (e.g. a full microprocessor), trimming at several temperatures is indeed acceptable, despite of the need to inform the many correction parameters for each sensor when programming the temperature-related software.

The major problem of such method, however, is the need of calibration at multiple temperatures too. This increases enormously the sensor production and test time and thus its cost.

[15]The trimming circuit also shares the same DEM circuitry (switches and part of the digital control), which allows a less complex front-end implementation.

[16]Although digital trimming can be implemented on chip, external trimming is much more flexible, since it is all time (re-)programmable.

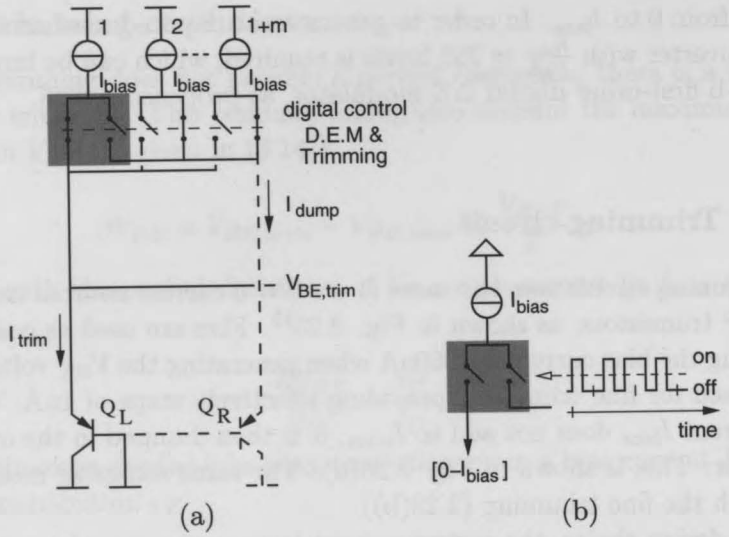


Figure 3.23: (a) V_{BE} trimming schematic. The digital logic used to program the trimmed bias current through the PNP transistors is also used for DEM control. The first current source is bitstream controlled. (b) Detail of the bitstream trimming.

3.10 Curvature correction

The base-emitter voltage V_{BE}, as discussed in Chapter 2, is not an ideal linear function of temperature. Many curvature correction techniques thus have been reported in the literature so to compensate for the non-linearity of V_{BE} which impairs the design of high precision temperature sensors.

Some of these curvature-correction techniques use non-PTAT currents or add a non-linear cancellation term to the bandgap voltage (or to V_{BE}). A good survey about all these methods are available either in [32] or [22].

A very common technique, probably because of its simplicity, consists in biasing the PNP transistors with a PTAT current. It is well-known that the curvature of V_{BE} of a transistor biased at a constant current ($m = 0$, refer to Chapter 2) is larger than if biased at a PTAT current ($m = 1$). This can be understood by realizing that, when the voltage V_{BE} start decaying faster then if linearly dependent on temperature, the bias current increases, compensating this decay to some extent. A bias current with the same temperature dependency of that of the saturation current would be better, but the complexity (and probably power) needed to generate such bias current and implementation issues in standard CMOS technologies are limiting factors of its applicability.

Yet, another way of correcting the curvature of V_{BE} is presented in [55]. In this situation, instead of having a temperature independent reference voltage V_{REF} (as given in (3.5)), a slightly temperature dependent reference voltage is used. With such temperature dependency, the non-linearity in the ratio μ due to the curvature in V_{BE} can be diminished (which turns into a smaller non-linearity in the temperature reading T_{dig} and error). Such correction, called ratiometric curvature correction [22] is possible because a linear temperature dependence in V_{REF} is also source of non-linearity in μ. This non-linearity, by proper design, can compensate for the second-order non-linearity due to curvature of V_{BE}. By precisely tuning the reference voltage V_{REF}, the temperature reading has no second-order non-linear component, but only a remaining (residual) third-order term. This allows the achievement of much more accurate temperature readings.

The implementation of the ratiometric curvature correction can be done by designing the factor α a bit larger than the value given in (3.7), i.e. 8% larger [53] or alternatively, like in [22], by slightly increasing the PNP biasing, given a constant factor α. Although a different knob, it has the same effect.

3.10.1 Curvature correction and trimming

Because trimming requires proper increase/decrease of the bias current, it can help finding the appropriate bias current so to compensate for the curvature in V_{BE}. This is obviously related to the size of the PNPs, and tuning is done taking into account the current density flowing through the bipolar transistors.

Since the design decision was to set the nominal trimming current to two unit current sources, a careful sizing of the PNP's was carried out so to obtain a strong non-linearity cancellation with a bias current close to 0.5μA (which corresponds to the second coarse trimming mode), as specified for proper and balanced trimming. This is shown in Fig. 3.24. For $trim = 2$, the error due to curvature in V_{BE} goes down from ± 0.6°C (with a quadratic characteristic) to something below ± 0.15°C (third-order characteristic).

Certainly, higher trimming modes would mean more relaxed trimming-resolution specifications, according to (3.81). However, higher modes would also mean more power consumption when generating the voltage V_{BE}, not to mention an unbalanced trimming capability. Figure 3.25(a) shows the main signals (V_{BE}, $\alpha \cdot \Delta V_{BE}$, V_{REF} and the bias current I_{bias}) for $trim = 2$. The smaller non-linearity in μ (and error), like mentioned, is related to the weak but finite dependency of V_{REF} with temperature (shown in detail in Fig. 3.25(b)).

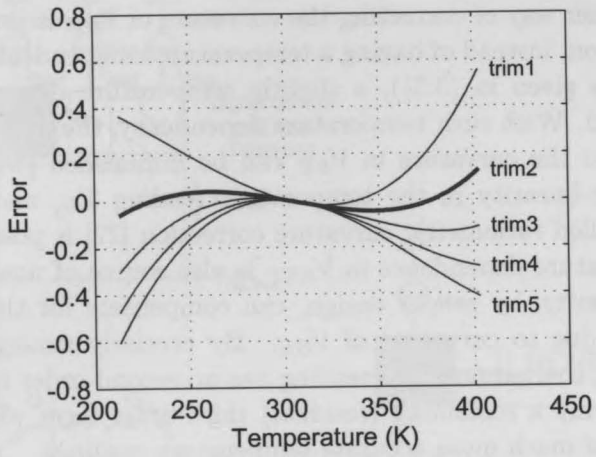


Figure 3.24: Non-linear temperature error for all five (coarse) trimming modes, i.e. from $trim = 1$ to 5 (bias currents ranging from 250nA to 1.25μA, in steps of 250nA). The PNP area was designed to give the smallest error for $trim = 2$, i.e. $I_{bias} = 0.5\mu$A (@ $T = 300$K).

3.11 Conclusions

The fundamental concepts for bandgap references were addressed as they provide the floor for the implementation of temperature sensors based on the ratiometric temperature measurement.

The current ratio m, the actual most relevant design parameter of the PTAT voltage ΔV_{BE}, was extensively analyzed. Errors on this current ratio were studied. The goal was to determine trade-offs so to be able to define it as well as the bias current given accuracy and power consumption specifications.

Since the bias current spread, as discussed, adds to V_{BE} spread, efforts to design an accurate bias current were described too. The β-compensation technique was presented in order to avoid the dependence of V_{BE} on β, further reducing the V_{BE} spread as well.

Based on the results obtained, the front-end implementation, comprised of a front-end core and a pre-bias circuit (PMOS current sources providing bias to the PNP transistors plus a low-power OTA), was described. Efforts to reduce the tail current of this amplifier were based on the reduction of its output load, the front-end current mirror (PMOS current sources).

A current ratio of $m = 5$ sets the best accuracy and power consumption trade-off. Interestingly, the current ratio p in the pre-bias set equal to the current ratio m in the core may help improving matching levels by operating

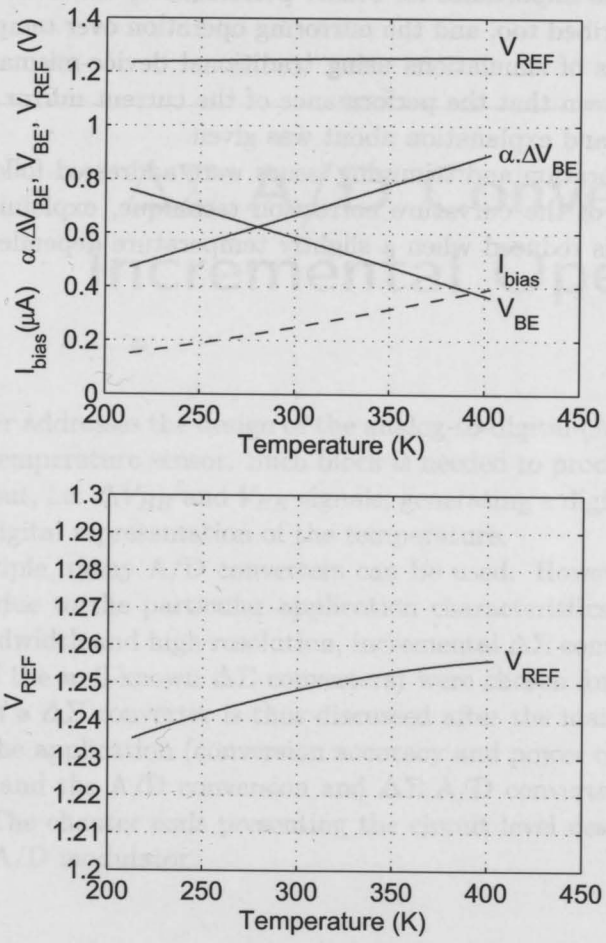


Figure 3.25: (a) Typical bandgap signals. For $I_{bias} = 0.5\mu$A, ($trim = 2$), the reference voltage has a small positive temperature dependency. (b) Reference voltage is shown in detail. Such slightly increase with temperature helps to cancel effects due to curvature of V_{BE}.

the pre-bias and core circuits under more similar conditions, i.e. under the same current ratios. A nominal bias current of 250nA at room temperature was needed given accuracy requirements. For a final inaccuracy smaller than 0.1°C, the total supply current found was below 5μA.

Because of its importance for sensor performance, the front-end current mirror was described too, and the mirroring operation over temperature analyzed by means of simulations using traditional device-mismatch models. Results have shown that the performance of the current mirror is temperature dependent and explanation about was given.

Finally, calibration and trimming issues were addressed followed by the the description of the curvature correction technique, explaining how the V_{BE} curvature is reduced when a slightly temperature dependent reference voltage is used.

CHAPTER 4

$\Delta\Sigma$ A/D Converters - Incremental Operation

This chapter addresses the design of the analog-to-digital (A/D) converter of the smart temperature sensor. Such block is needed to process the front-end analog output, i.e. ΔV_{BE} and V_{BE} signals, generating a digital signal, whose value is a digital representation of the temperature.

In principle, many A/D converters can be used. However, as it will be presented, due to the particular application characteristics and needs, e.g. narrow bandwidth and high resolution, incremental $\Delta\Sigma$ converters (a particular case of the well-known $\Delta\Sigma$ converters) were chosen for this work. The operation of a $\Delta\Sigma$ converter is thus discussed after the main design aspects related to the application (conversion accuracy and power consumption) are introduced and the A/D conversion and $\Delta\Sigma$ A/D converters fundamentals reviewed. The chapter ends presenting the circuit level design back-ground of the $\Delta\Sigma$ A/D modulator.

4.1 Introduction

In recent CMOS technologies, the speed, accuracy and power-consumption trade-off imposed by mismatch considerations is about two orders of magnitude more demanding than the relationship imposed by thermal noise considerations [56][11]. This trade-off is expected also in $\Delta\Sigma$ A/D converters, perhaps confirming that one of the most important reasons to use oversampling and noise shaping methods is the well-known relative insensitivity of these techniques to the mismatch between components, which is quite a severe problem in modern technologies [56].

Despite, however, of the relaxed demands on matching, to achieve a certain bandwidth and resolution with an L^{th}-order $\Delta\Sigma$ converter, it is always necessary to cope both with the thermal and quantization noise. Since the thermal noise decreases 0.5 bit for each doubling of the sampling frequency, while the quantization noise $L+0.5$ bit for the same frequency increase [56], the main limitation of an A/D converter would be in principle set by the thermal noise which can only be decreased at the cost of capacitor area (larger capacitors) and thus power consumption (due to larger loads). However, in the case of precision smart temperature sensors, other errors have to be taken into consideration. This is further investigated on the basis of the needed requirements, as follows.

4.1.1 Requirements

The overall sensor accuracy and power consumption depends not only on the front-end that generates ΔV_{BE} and V_{BE} signals but also on the A/D converter that converts these signals to a digital format, providing a digital reading of temperature. Thus, it is very important to design it without adding conversion errors. The accuracy and power requirements used to derive specifications for the front-end are valid for the A/D converter too.

Accuracy

The specified temperature inaccuracy should be below than $\pm 0.1°C$ over the entire the military range, i.e. from $-55°C$ to $125°C$. This sets the overall inaccuracy for the A/D converter to below $\pm 0.1°C$, since the front-end also contributes to this quantity. Taking into account the various sources of errors in the A/D conversion, e.g. sampling errors, offset voltages, errors due to non complete settling or charge injection, it is acceptable the need of having them canceled or reduced to negligible values, i.e. $\pm 0.01°C$. However, in order to have a low-power conversion though, the amplitude of such errors are allowed to increase, provided the final specified accuracy is achieved.

Power consumption

In an ideal scenario, the power consumption of the A/D converter would be negligible compared with that of the front-end, so the power consumed by the sensor would be actually very close to that required by the circuit that generates the temperature information. However, accurate data conversion is well known to dissipate power, even if at low conversion speed. Taking this into account, the approach of this work was to design an A/D converter able

to meet the accuracy requirements (e.g. as good as the state-of-the-art accuracy over the same military range), while optimizing for power consumption. In practice and based on the prior-art knowledge and technical reviews, such approach resulted in the following specification for the modulator: power consumption comparable to or at worse few times higher than the power consumed in the front-end, i.e. current budget for the modulator of or below approximately 15μA (for a supply voltage varying from 2.5 to 5.5V).

Resolution and effective number of bits

The resolution required at the output of the A/D converter is set by the temperature sensor. Because an inaccuracy below $\pm 0.1°$C is needed, a resolution of $\pm 0.01°$C might be recommended. In this case, the quantization noise error can be neglected when compared to the accuracy requirement. However, often a higher resolution means more energy for a conversion. It is important to analyze if lower resolutions are accepted, i.e. if a rise in the quantization noise does not prevent the accuracy specifications to be met over the entire temperature range of operation.

When working with resolution, the Effective Number of Bits (ENOB) is often adopted. A resolution of $\pm 0.01°$C, i.e. a Least Significant Bit (LSB) of $0.02°$C, over a temperature range of $600°$C will result in $\frac{600}{0.02} = 30,000$ steps, an approximate ENOB of 14.8 bits.

Bandwidth

The speed of conversion or conversion time T_C is the time needed for the converter to complete a single conversion and generally fixes the maximum allowable input sampling rate f_s, i.e. $f_s = \frac{1}{T_C}$. The maximum number of conversions per second, respecting the maximum sampling rate is known as sensor bandwidth [57]. As introduced in the first chapter, a temperature sensor with a bandwidth of 10Hz is enough in most of the applications for monitoring because of the thermal inertia and other thermal properties of the integrated sensor package [22]. Application specific temperature sensors, e.g. those for control systems, might be required to have a larger bandwidth.

4.1.2 A/D converter architectures

In a smart temperature sensor, an A/D converter is required to provide the temperature in a digital format, readily readable by most of digital systems, e.g. microprocessor and micro-controllers. Which A/D architecture however, depends on the application and on the particular specifications that have to be met.

In ultra-low power applications, e.g. autonomous (sensor) systems, sensor networks [5] and other battery-supplied systems where there is a serious lack of power for operation, low-power consumption is the most important requirement and not accuracy (that comes often in second). In these applications, sensors are required to dissipate low power in order to preserve the system's operating lifetime. Total quiescent currents of 50μA [58] and below are required.

Near-Nyquist converters like Successive Approximation Register (SAR) A/D converters are the most indicated architecture [5] since they undoubtedly provide the lowest power dissipation when compared to oversampled converters. But such architectures have their accuracy strongly dependent on capacitor matching (which are used to build the internal D/A converter such architecture needs). That is why most of the designs based on SAR architecture are often limited to 10 to 12 bits of resolution. Another issue concerns the overall system's power consumption. If the entire sensor is considered, the total power dissipation may not necessarily be smaller, especially if more than 12 bits of resolution is required. SAR architectures are matching limited, and higher resolutions are attainable with larger devices and/or expensive auto-calibration and trimming procedures [59][60], which increases the area and power consumption. The anti-alias filter implementation complexity has to be accounted for too [5].

If the application has a narrow bandwidth, ΔΣ A/D converters have proved to be the best choice. These converters exchange resolution in time for that in amplitude by combining sampling at well above the Nyquist rate with coarse quantization embedded within a feedback loop in order to suppress the amount of quantization noise appearing in the signal band [61], not relying then on matching. Other advantage is higher immunity against interferences due to the oversampled operation. Despite of mixed solutions, where ΔΣ A/D converters are combined together with other architectures sometimes referred as two-step converters [60], Extended-Counting [62] or Zoom A/D converter [13] to get higher resolution and more energy-efficient operation, higher orders ΔΣ A/D converters have also been suggested with the same advantages [63][64].

4.2 Analog-to-digital (A/D) conversion

The process by which an analog signal is encoded into a digital representation encompasses both the sampling of an analog waveform in time and the encoding or quantization of it in amplitude. The minimum rate at which a signal can be sampled is determined by its bandwidth f_B, while the resolu-

tion with which it is encoded establishes the amount of quantization error e_q that can be tolerated in the conversion [61].

4.2.1 Sampling and quantization

Analog signals are continuous both in time and amplitude and their spectra are limited in frequency (extending from DC to f_B, i.e. the signal bandwidth) as an effect of their continuity in amplitude. In this way, the sampling operation is required to sample and hold a constant signal at the input of the A/D converter while the amplitude is being quantized (encoded) to a given level.

The anti-alias filter is used to limit the bandwidth of the analog signal to less than a half of the sampling frequency f_S, thereby ensuring that the sampling operation will not alias noise or out-of-band signal into the base band according to the Nyquist theorem, i.e. $f_S \geq 2f_B$.

Next to the sampling comes the quantization operation. A quantizer encodes a sampled continuous-amplitude signal into a set of discrete levels. A quantizer that maps its input range into 2^n output levels is referred to as an n-bit quantizer. In a uniform quantizer, the input signal x is divided into segments of equal width, each of which maps to a single output code. The separation between the input levels is

$$\delta = \frac{V_{FS}}{2^n} \tag{4.1}$$

where V_{FS} is the so called full-scale input range.

Since the quantizer resolution is not infinite, it implies an error known as quantization error and defined as the difference between the analog input x and digital output y, i.e,

$$e_q = y - x, \tag{4.2}$$

once y is reconfigured in the analog domain [57]. Such error, appearing as a sawtooth shaped waveform (for a ramp function as input), is inherent of any A/D converter. As long as the input x lies within the input range, the magnitude of the quantization error of an ideal converter is limited to $\pm\delta/2$ [65].

While the average of the quantization error e_q is zero if this error is assumed having equal probability $P(e_q)$ of lying anywhere in the range $\pm\delta/2$ (i.e. uniformly distributed from $-\delta/2$ to $+\delta/2$ with $P(e_q) = 1/\delta$), it has though a mean square value (quantization error power) given by

$$e_{q,rms}^2 = \int_{-\delta/2}^{+\delta/2} e_q^2 \rho(e_q) \mathrm{d}e_q = \frac{\delta^2}{12}. \tag{4.3}$$

The Signal-to-Noise Ratio (SNR) of a n-bit quantizer is then defined as the power of a full-scale sinusoid[1] divided by the quantization error power:

$$SNR = \frac{\frac{\delta^2}{8}2^{2n}}{\frac{\delta^2}{12}} = \frac{3}{2}2^{2n}. \tag{4.4}$$

or, when expressed in dB, by

$$SNR = 6.02n + 1.76 \text{ [dB]}. \tag{4.5}$$

This well-known expression relates the SNR of an A/D converter to the number of bits or resolution of its quantizer. It gives the best SNR for an n-bit A/D converter [60]. Nonetheless, depending on the input signal, i.e. for a reduced input signal, the SNR can be lower, case when the number of bits n is named ENOB. Equation (4.5) shows that 1 bit increase in resolution corresponds to a 6dB increase in SNR. However, while in Nyquist converters the SNR of a converter is equal to the DR of its quantizer, in oversampling A/D converters, the DR of the converter can be much greater than that of its quantizer [61].

4.2.2 Oversampling

Oversampling can be used to lower the quantization noise, thus increasing the SNR and consequently resolution. It occurs whenever a signal is sampled at a frequency larger than twice its bandwidth f_B. According to the Nyquist theorem, no extra information is obtained, since the spectrum is only widened. But it is exactly this widening the key advantage of oversampling, since in this case the signal band occupies a smaller fraction of the Nyquist interval, making possible to use digital cancellation on the relatively large fraction of the quantization noise that is outside the band of interest [66]. In other words, because the quantization error power, given by (4.3), remains unchanged regardless of the sampling rate, the quantization error power density must go down as the spectrum widens [57]. That is, for $f_S = \frac{1}{T}$, the entire quantization error $e^2_{q,rms}$ equally distributed from 0 to $f_S/2$ establishes the power spectral density of the sampled noise $S^2_{e_q}(f)$[2]

$$S^2_{e_q}(f) = \frac{\frac{\delta^2}{12}}{f_S/2} = \frac{\delta^2}{12}\frac{1}{f_S/2}. \tag{4.6}$$

[1] $\left(\frac{V_{FS}}{2\sqrt{2}}\right)^2 = \frac{\delta^2 2^{2n}}{8}$

[2] The power is assumed entirely in the right side of the frequency axis.

This shows that the higher f_S is, the lower the $S^2_{e_q}(f)$ is. In the case of a Nyquist-rate converter, the total in-band noise power is exactly the same quantization noise as given in (4.3), since f_B coincides with $f_S/2$. In the case of an oversampled converter, where the sampling frequency is OSR (Oversampling Rate) times higher than $2f_B$, i.e.

$$OSR = \frac{f_S/2}{f_B}, \tag{4.7}$$

if we restrict the bandwidth of the oversampled signal to f_B, a lower noise power will be present inside the signal band given by

$$n_B^2 = \int_0^{f_B} \frac{\delta^2}{12}\frac{1}{f_S/2}\mathrm{d}f = \frac{\delta^2}{12}\frac{1}{f_S/2}f_B = \frac{e^2_{q,rms}}{OSR}. \tag{4.8}$$

That is, oversampling reduces the in-band rms noise from ordinary quantization by the square root of the oversampling ratio [65]. In this case, the SNR, formerly calculated in (4.4) is now given by

$$SNR = \frac{3}{2}OSR2^{2n} \tag{4.9}$$

or, in dB, as

$$SNR = 6.02n + 1.76 + 3\log_2(OSR)\ [\mathrm{dB}]. \tag{4.10}$$

The SNR improvement resulting from oversampling is graphically shown in Fig. 4.1(a) and (b) by comparing the shaded areas. With $OSR = 4$, the noise power density is four times lower (so four times lower is the total power that falls into the signal band). This improves the SNR by 6dB, which is the same of adding one bit to the quantized signal. Although it is possible to arbitrarily increase the number of bits, this is at the expense of a rapidly growing OSR, which severely restricts the system bandwidth, and complicates the system design due to increased frequency operation (not to mention the power consumption). To improve the SNR without compromising the bandwidth, oversampling alone is not sufficient. A technique called noise shaping is often used.

4.2.3 Noise shaping

Noise shaping is used to further increase the conversion performance by high-pass filtering the quantization noise, displacing most of its power from low frequencies where the input signal is placed to higher frequencies close to $f_S/2$ [67]. This is shown in Fig. 4.1(c). The actual remaining quantization

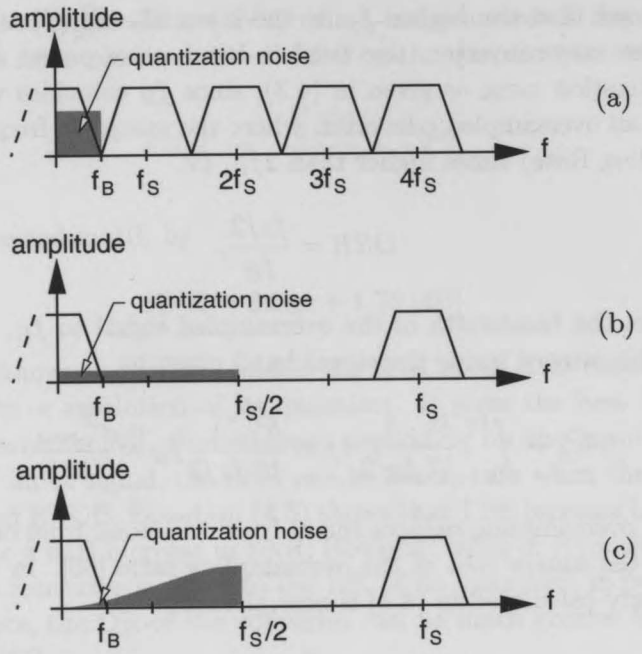


Figure 4.1: Signal spectrum of a (a) Nyquist-rate converter, (b) oversampled converter and (c) oversampled with noise shaping converter.

noise power inside the signal band depends on the type of filtering applied, i.e. filter architecture and order and cut-off frequency.

Noise shaping or filtering is achieved by putting a feedback loop around the quantizer as shown in Fig. 4.2. Probably the most known method of implementing such feedback is the ΔΣ loop or modulator, as depicted, with the loop filter setting the noise shaping characteristic. The input to the circuit feeds to the quantizer via an integrator, and the quantized output feeds back to subtract from the input signal. This feedback forces the average value of the quantized signal to track the average input. Any persistent difference between them accumulates in the integrator and eventually corrects itself [68][65]. In other words, the loop controls the quantizer input in such a way that the output tracks the input signal as close as possible.

The loop though cannot affect the heights of the steps the quantizer outputs, but it tends to modulate the steps of the quantized signal to try to average out the difference between the x and y signal [57]. Since the quantized signal is oversampled, the output looks like a high frequency digital noise superposed on the input signal. This increases the high frequency noise of the quantized signal, but lowers the low frequency components.

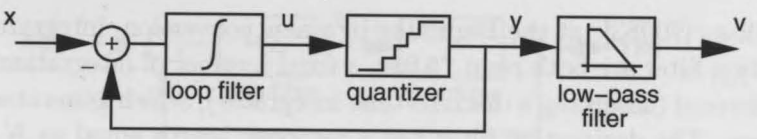


Figure 4.2: $\Delta\Sigma$ A/D converter. Noise shaping is implemented by a $\Delta\Sigma$ modulator. Low-pass filtering is needed to filter the high frequency noise at the quantizer output out.

4.2.4 $\Delta\Sigma$ A/D converters

The loop filter of Fig. 4.2 provides gain while it determines the bandwidth of the noise shaper. At low frequency, where the gain is high, the quantization noise is almost canceled. However, at high frequencies where the loop gain is low, the noise is increased. However, this happens outside the signal band f_B. With a (digital) low-pass filter after the modulator (see Fig. 4.2), that ideally cuts all frequencies over f_B, the noise excess is thus filtered out. Moreover, the low-pass filter also decimates the modulator output, the bitstream y, providing a digital word as output. Such structure composed of a $\Delta\Sigma$ modulator and low-pass decimation filter composes the well-known $\Delta\Sigma$ A/D converter.

4.2.5 Incremental operation mode

Analog to digital converters used in instrumentation and measurement applications (e.g. temperature sensors in the particular case of this work) often require very high absolute accuracy and linearity, and very low offset and noise. Low-power is also an important consideration. On the other hand, the frequency band of the input signal is usually very narrow, a few hertz wide [69]. Such specifications are fulfilled by incremental $\Delta\Sigma$ data converters [70] due to its characteristics: very precise conversion with accurate gain, high linearity and low offset [64]. For higher-order structures, the conversion time can be relatively short [63][56].

Incremental data converters, though, operate differently from the traditional $\Delta\Sigma$ A/D converters, in the so called transient mode (instead of operating continuously) where the A/D converter converts an individual input sample at each time. Moreover, the input signal is usually a DC or a near-to-DC signal, instead of an AC signal[3]. In general words, its operation

[3]Despite of the different operation, incremental $\Delta\Sigma$ converters are structurally similar to the conventional $\Delta\Sigma$ converters, sharing thus in general terms, the same classical theory of $\Delta\Sigma$ converters.

is as follows [70][63]: at the beginning of a new conversion, integrator(s) and decimation filter are both reset. After, a fixed number of integration steps N are performed (assuming a discrete-time integrator), which generates a N-bit bitstream. The decimation filter has a response length equal to N used for down-sampling with a decimation factor N.

Since for each sample, N bits are generated, the number of cycles N has a similar interpretation of oversampling ratio in classical ΔΣ converters. The exact number of cycles N though will depend on the required resolution in bits which is set by the application but also on some particular A/D converter characteristics, e.g. modulator topology and order, decimation filter, etc.

4.3 ΔΣ modulators

A ΔΣ modulator (with a typical diagram shown in Fig. 4.3(a)) requires the implementation of two transfer functions. In the case of a temperature sensor, these functions are a low-pass Signal Transfer Function (STF) and a high-pass Noise Transfer Function (NTF), both implemented with an architecture containing a loop filter $H(z)$ and a coarse quantizer. Fig. 4.3(b) presents the ΔΣ modulator with the quantizer replaced by its linear model. In this model $E(z)$ is an additive, uniformly distributed quantization error and uncorrelated with the modulator input (the same characteristic of a white noise). Since the quantizer is a non-linear circuit, and the error can have strong correlation with the input, the model has limited applicability, and time-domain simulations are often necessary to verify the converter operation. Nonetheless, the analytical analysis brings insight and good understanding of the ΔΣ operation.

4.3.1 ΔΣ modulators with discrete-time loop filters

Consider the typical diagram of a ΔΣ modulator as shown in Fig. 4.3(a). The modulator input and output in the z-domain are given by $X(z)$ and $Y(z)$, respectively while the signal $E(z)$ models the quantization error. For this system, the following equation relating the output $Y(z)$, input $X(z)$ and noise source $E(z)$ can be written

$$Y(z) = [X(z) - Y(z)]H(z) + E(z). \tag{4.11}$$

or, after isolating $Y(z)$

$$Y(z) = \frac{H(z)}{1 + H(z)}X(z) + \frac{1}{1 + H(z)}E(z) \tag{4.12}$$

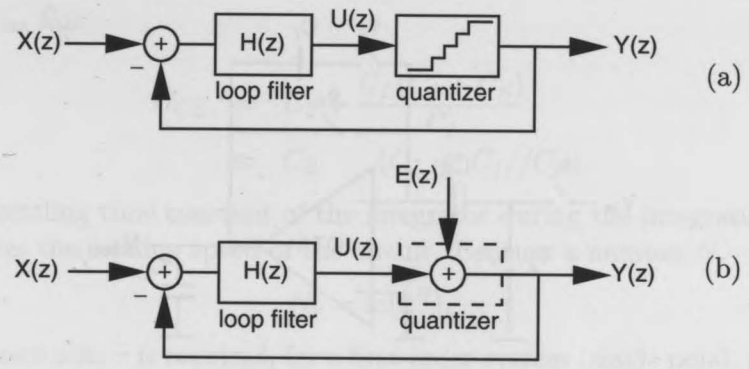


Figure 4.3: (a) ΔΣ modulator and (b) ΔΣ modulator with the quantizer replaced by its linear model.

Assuming that both inputs $X(z)$ and $E(z)$ are not correlated, the output of the modulator can be written as

$$Y(z) = STF(z)X(z) + NTF(z)E(z) \tag{4.13}$$

with the Signal Transfer Function (STF) given by

$$STF(z) = \frac{Y(z)}{X(z)} = \frac{H(z)}{1+H(z)}. \tag{4.14}$$

and the Noise Transfer Function (NTF) given by

$$NTF(z) = \frac{Y(z)}{E(z)} = \frac{1}{1+H(z)}. \tag{4.15}$$

The $STF(z)$ is expected to be one at low frequency, while the $NTF(z)$ zero, if the filter gain in the baseband is high enough. In this way, the modulator output replicates the input, and the amount of noise power remaining in the baseband is thus lowered. The $STF(z)$ expresses how the input signal is filtered by the modulator, but because this is a linear (distortion less) operation [71], the $STF(z)$ is not critical. In this way, is the $NTF(z)$ the defining parameter of a ΔΣ modulator, since all major characteristics like performance and stability are set by the $NTF(z)$. Both functions can be arbitrarily synthesized. However, not all $NTF(z)$s are realizable, and if realizable, not necessarily will give stable modulators.

4.3.2 Current estimation of a ΔΣ modulator

In this section the supply current of a ΔΣ modulator is estimated to find the modulator's power consumption given a supply voltage (or the energy per

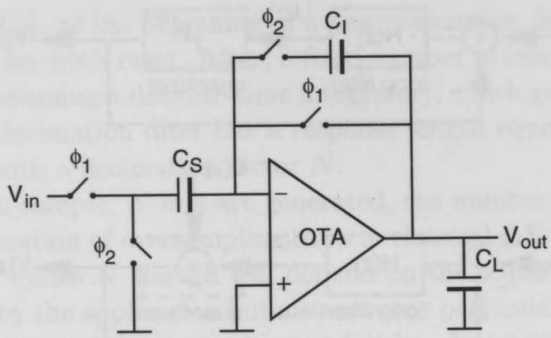


Figure 4.4: Schematic of a switched-capacitor integrator. Phases $\phi 1$ and $\phi 2$ are non-overlapping.

conversion, given a conversion time).

The supply current of a $\Delta\Sigma$ modulator is mainly given by the current consumed by the integrators and by the comparator. Since at least one comparator is needed, regardless of the modulator order, its analysis is skipped here, but it has to be accounted for when determining the total modulator supply current.

Figure 4.4 shows a simplified schematic of a switched capacitor integrator. The transconductance g_m of the operational transconductance amplifier OTA, used to implement the integrator, serves as a current consumption model for the integrator [67]. It is given by

$$g_m = \frac{1}{\tau} C_{CL} \tag{4.16}$$

where τ is the integrator settling time constant, when in closed loop, and C_{CL} the closed-loop capacitance during the integration phase. The closed-loop capacitance C_{CL} is given by the open-loop capacitance C_{OL}

$$C_{OL} = C_L + \frac{C_I C_S}{C_I + C_S} \tag{4.17}$$

$$\approx \frac{C_I C_S}{C_I + C_S} \qquad (C_L \ll C_I // C_S) \tag{4.18}$$

where C_S, C_L, C_I are the sampling, load and integration capacitors, respectively, divided by the (e.g. non-inverting) feedback factor β [72] (refer also to Appendix F)

$$\beta = \frac{C_I}{C_I + C_S} \approx \frac{C_I}{C_I + C_S}, \tag{4.19}$$

i.e. $C_{CL} = \frac{C_{OL}}{\beta}$:

$$C_{CL} = C_S + \frac{C_L(C_I + C_S)}{C_I} \tag{4.20}$$

$$\approx C_S \qquad (C_L \ll C_I // C_S). \tag{4.21}$$

The settling time constant of the integrator during the integration cycle determines the settling speed of the circuit. Because a number N_τ

$$N_\tau = \ln(2^B), \tag{4.22}$$

of time constants τ is required, for a first-order system (single pole), to settle to a full precision (B bits) [56], it results that

$$\tau = \frac{1}{N_\tau}\frac{T_S}{2} = \frac{1}{\ln(2^B)2f_S} \tag{4.23}$$

since half of the clock period T_S is usually available for settling[4]. Combining (4.16) and (4.23) finally results in

$$g_m \approx \ln(2^B)2f_S C_S. \tag{4.24}$$

For a CMOS OTA implementation, the transconductance of the integrator is directly related to the transconductance of the MOS transistor in the input differential pair given by

$$\frac{g_m}{I_D} = \frac{2}{V_{eff}} \tag{4.25}$$

which results, if solved for I_D, in

$$I_D = g_m \frac{V_{eff}}{2} \tag{4.26}$$

where I_D is the transistor biasing current, and V_{eff} the effective transistor voltage, which depends on the transistor region of operation.

Considering that the total OTA current is given by K_{OTA} times the bias current I_D, and that higher order integrators have bias currents that can be progressively scaled down relatively to the first integrator, and that the sum of all scaling factors are given by $K_{\Sigma\Delta}$, the total current consumed by the integrators can be estimated by combining (4.24) and (4.26):

$$I_{total} = K_{\Delta\Sigma} K_{OTA} \left[\ln(2^B) f_S C_S V_{eff}\right]. \tag{4.27}$$

Based on the estimation of the total current of a modulator, given by (4.27), the following trade-offs can be a derived, which helps to understand the power consumption in a $\Delta\Sigma$ modulator:

[4]More accurate modeling has to take into account the slewing time [67].

1. Use low-order modulators, i.e. smaller $K_{\Delta\Sigma}$: however, since in a $\Delta\Sigma$ converter the power consumption is mainly due to the first integrator, additional integrators often do not contribute significantly ($K_{\Delta\Sigma} \approx 2$ is observed for high-order $\Delta\Sigma$ modulators [56][5], e.g. [61]). Besides, low-order modulators require more cycles for a given resolution, which means a higher frequency f_s for the same conversion time, suggesting that higher-order modulators have advantages;

2. Use topologies where K_{OTA} is smaller, e.g. telescopic- versus folded-cascode, if a reduced output signal swing can be tolerated;

3. Minimize V_{eff}, e.g. bias the OTA input pair in weak inversion;

4. Minimize the sampling capacitor C_S: however, the sampling capacitor cannot be reduced arbitrarily, since it is set by thermal noise constraints as given by

$$N_{th} = \frac{1}{OSR}\frac{kT}{C_S} \tag{4.28}$$

   and other requirements, e.g. matching requirements. Still, designing the sampling capacitor slightly higher (17%) than the minimum size may even give a less total power consumption [50];

5. Design for the minimum resolution the application needs;

6. Minimize f_S: however, this is often prevented by resolution and conversion time requirements.

In practice, in face of many underlying trade-offs, design decisions are taken to satisfy the most relevant application requirements, e.g. accuracy and power consumption.

4.3.3 Charge balancing

Figure 4.5(a) shows the block diagram of a first-order $\Delta\Sigma$ modulator. Every clock cycle, a charge is accumulated due to the integration of the voltage present at the integrator's input

$$V_{in} - bsV_{REF} \tag{4.29}$$

[5]This suggests that the first modulator typically consumes half of the total modulator's power consumption in high-order loops. In the case, e.g. of a second-order modulator, $K_{\Delta\Sigma} \approx 1.5$ is expected.

where V_{in} is the input signal, V_{REF} is the reference voltage and bs is the bitstream value (comparator output of the former ΔΣ cycle). Depending on bs, either a voltage V_{in}

$$V_{in} = \alpha \Delta V_{BE}, \tag{4.30}$$

when $bs = 0$, or a voltage $V_{in} - V_{REF}$

$$V_{in} - V_{REF} = \alpha \Delta V_{BE} - (V_{BE} + \alpha \Delta V_{BE}) = -V_{BE}. \tag{4.31}$$

when $bs = 1$, is present at the integrator's input. This operation, when the charge accumulated due to the integration of V_{in} is eventually balanced by the charge accumulated due to the integration of V_{in} minus V_{REF}, is commonly referred as charge balancing [22].

Since $V_{in} = \alpha \Delta V_{BE}$ and $V_{REF} = \alpha \Delta V_{BE} + V_{BE}$, an equivalent but more insightful ΔΣ modulator block diagram is shown in Fig. 4.5(b), where the front-end's output signals are emphasized as the modulator input. A simpler and perhaps more efficient implementation though, introduced in [22], is shown in Fig. 4.5(c). This type of implementation can be understood realizing that the integrator's input, depending on bs, is either the value given in (4.30) or the value given in (4.31), and therefore has the same function of an analog multiplex, what is exactly represented in the picture. The main advantage of such implementation is that no explicit V_{REF} is needed, since it is built internally.

Because of the modulator's feedback, regardless of the implementation type, the average at the input of the integrator over time must be zero, i.e.

$$\frac{(1-\mu)TV_{in} + \mu T\,(V_{in} - V_{REF})}{T} = 0 \tag{4.32}$$

which results, if solved for μ, in

$$\mu = \frac{V_{in}}{V_{REF}} = \frac{\alpha \Delta V_{BE}}{\alpha \Delta V_{BE} + V_{BE}} \tag{4.33}$$

which is actually how the temperature is measured (refer to ratiometric temperature measurement in Chapter 3).

4.3.4 Dynamic range extension

Original V_{PTAT} voltage $\alpha \Delta V_{BE}$ (as shown in Fig. 4.6) covers a temperature range from 0 to 600K while only about 200K is required (i.e. from 218K to 398K). While the transfer function μ given in (4.33) is easily implemented, it uses roughly only 30% of the available input dynamic range of the A/D converter, since the extremes of the needed operating range correspond to μ ranging approximately from 0.33 to 0.67.

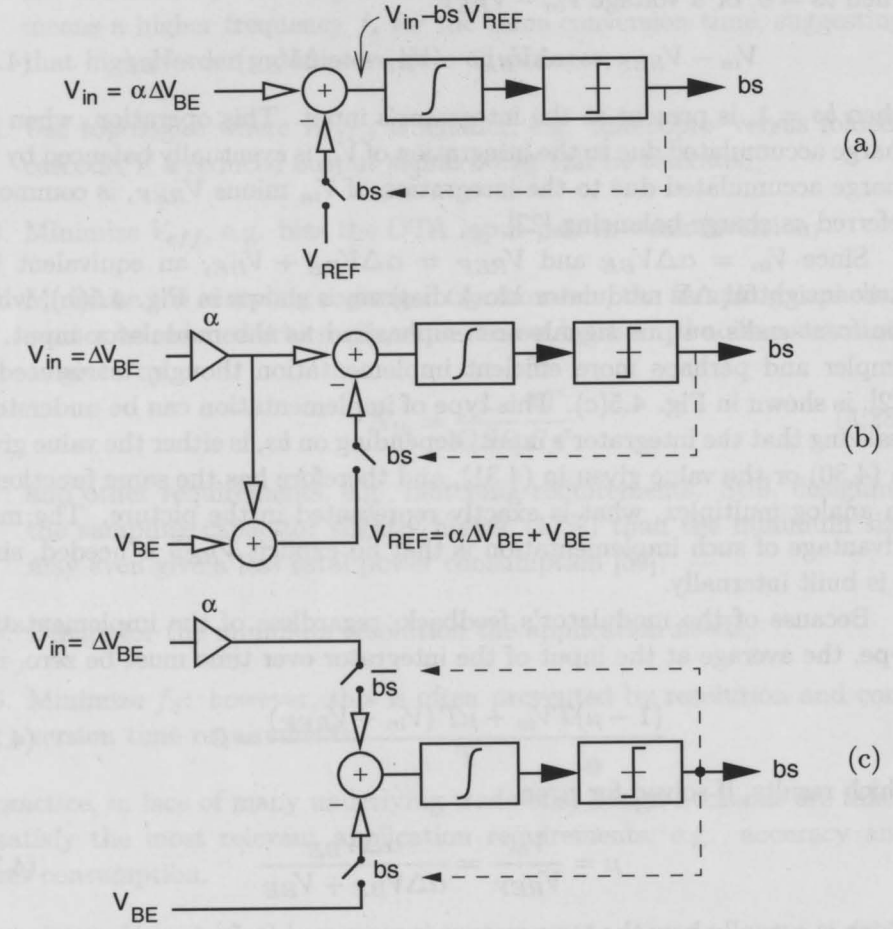


Figure 4.5: (a) Block diagram of a first-order ΔΣ modulator, (b) a direct implementation of the modulator emphasizing the front-end's output signals and (c) a simpler and more efficient implementation (with no explicit reference needed).

Table 4.1: Relation between the input signal range, number of bits (ENOB) for ±0.01°C resolution ($LSB = 0.02$°C) and number of conversion cycles (refer to Fig. 4.13) for a second-order ΔΣ modulator.

full-scale temp. range (K)	used input range	required ENOB	estimated # ΔΣ cycles	#cycles reduction
0-600	180/600 (30%)	14.87b	245	0 %
150-450	180/300 (60%)	13.87b	173	30%
200-400	180/200 (90%)	13.28 b	141	43 %

Since a smaller resolution (for the same accuracy) is required for the ΔΣ A/D converter if the input signal uses more of the available converter's input dynamic range, where a lower resolution often translates to less power consumption per conversion (once the conversion takes less ΔΣ cycles), a more efficient use of such dynamic range, at the expense of a more complicated combination of V_{BE} and ΔV_{BE}, has been proposed. First in [23] and latter in [24] (refer also to [73]). The proposed new voltage, sometimes named V_{TEMP}, can achieve percentages up to 90% (as shown in Fig. 4.6).

Although very attractive and efficient, the V_{TEMP} voltage as originally proposed is in general indicated for first-order ΔΣ converters. This is because the implementation of such voltage can be difficult in single-loop second-order (or higher-orders) ΔΣ modulators. In such modulators, the available input range is limited to 90% due to stability issues (while in practice, only 75% is suggested in designs) [63]. A (new) V_{TEMP} that uses less than 90%, e.g. 60% of the modulator's input range would be thus more appropriate.

Table 4.1 compares the former V_{PTAT} with V_{TEMP} of [24] and that proposed in this work (further addressed in this work) in the case of a second-order ΔΣ modulator in terms of ΔΣ cycles required given the specified resolution (which is a function of the used input voltage). All voltages are shown in Fig. 4.6. Although presenting some differences, the goal is the same: to modify the original V_{PTAT} voltage to obtain the same conversion accuracy within less ΔΣ cycles, which effectively might represent power savings.

4.3.4.1 Composition of V_{TEMP}

Classical V_{PTAT} is simply a scaled voltage of ΔV_{BE}, i.e. $\alpha \Delta V_{BE}$. Because the reference voltage V_{REF} is partially build from this voltage, a very clever charge balancing, implemented by multiplexing either $\alpha \Delta V_{BE}$ or V_{BE} as input to the A/D converter, is possible. The simplicity has price: it uses

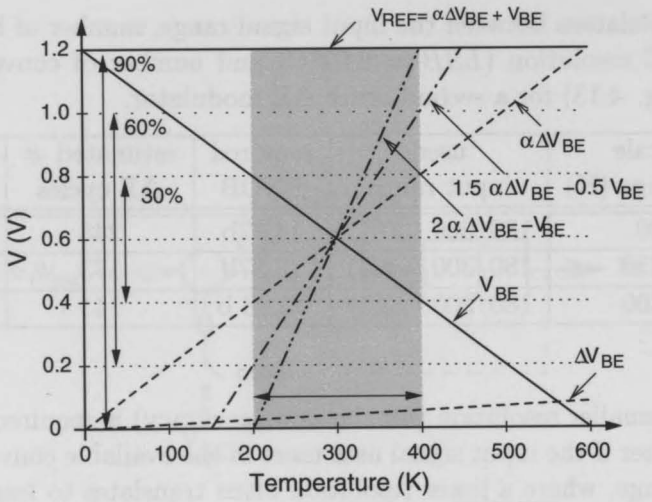


Figure 4.6: Original V_{PTAT}, V_{TEMP} as in [24] and here proposed mid-term V_{TEMP}.

only 30% of the input range of the A/D converter. The output μ is given by

$$\mu = \frac{\alpha \Delta V_{BE}}{V_{BE} + \alpha \Delta V_{BE}}. \tag{4.34}$$

The implementation of the new V_{TEMP} is based on same $\alpha \Delta V_{BE}$, but combines not only this voltage but also V_{BE} in such a way to define a different input signal, so more of the available input range is used. Several mathematical solutions are possible but practical aspects related to the physical implementation put severe restrictions on the design of most of them. Generally, the new V_{TEMP} can be written as

$$V_{TEMP} = (G - F)\alpha \Delta V_{BE} - F V_{BE} \tag{4.35}$$

where F and G are two gains used to establish a linear combination of both voltages. It is worth to mention that for $F = 0$ and $G = 1$, the new voltage is in fact the classical V_{PTAT}. For arbitrary F and G gains, the ratio μ of this new voltage with respect to the reference voltage is now given by

$$\mu = \frac{(G - F)\alpha \Delta V_{BE} - F V_{BE}}{V_{BE} + \alpha \Delta V_{BE}} \tag{4.36}$$

$$= \frac{G\alpha \Delta V_{BE} - F(V_{BE} + \alpha \Delta V_{BE})}{V_{BE} + \alpha . \Delta V_{BE}} \tag{4.37}$$

$$= \frac{G\alpha \Delta V_{BE}}{V_{BE} + \alpha \Delta V_{BE}} - F \tag{4.38}$$

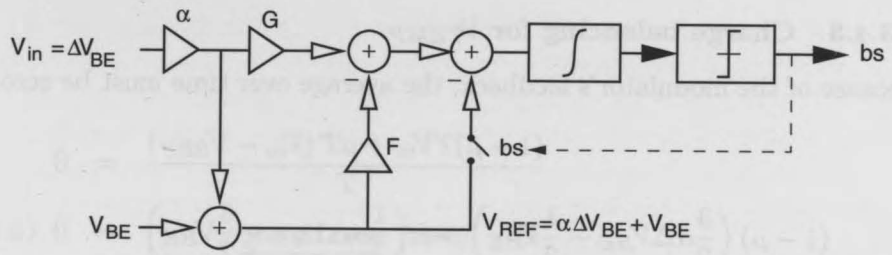


Figure 4.7: Charge-balancing model used to generate the new V_{TEMP}.

that is an augmented version of the former V_{PTAT} minus a constant, thus also a representation of the temperature.

The voltage V_{TEMP} is clearly more complex to implement (refer either to (4.35) or to Fig. 4.7). However, the same multiplexing used with the classical method (so to build the reference voltage internally) can be used here (not shown).

4.3.4.2 An optimized voltage V_{TEMP}

While many solutions are possible for F and G, given a factor α, in order to cover 60% of the input range with $\alpha = 16$[6], i.e. from 0.2 to $0.8V_{REF}$ with margin for trimming as well for the modulator operation (with no overload), F and G are set to $\frac{1}{2}$ and 2.0, respectively[7]. The new voltage V_{TEMP}, given then by

$$V_{TEMP} = \frac{3}{2}\alpha\Delta V_{BE} - \frac{1}{2}V_{BE}, \tag{4.39}$$

is shown in Fig. 4.6 and has the following main characteristics:

1. it uses 60% of A/D converter input range, compared with the former 30%, but it is more complex due to the other two gains that need to be implemented;

2. it is fully compatible with the front-end architecture, since it uses the same voltages $\alpha\Delta V_{BE}$ and V_{BE}, although with a different scaling;

3. it is now sensitive to process spread, since the voltage V_{BE} depends on the PNP saturation current I_S. This, however, leads to the same spread in μ because of presence of V_{BE} in V_{REF}.

[6]The factor $\alpha = 16$ is needed for a current ratio $m = 5$, as previously explained in Chapter 2.

[7]While simulation results for $F = 0.5$ and $G = 2.25$ have shown the best performance for a second-order $\Delta\Sigma$ modulator, the required $\Delta\Sigma$ cycle length (when compared with that of the PTAT voltage) was unacceptably long.

4.3.4.3 Charge balancing for V_{TEMP}

Because of the modulator's feedback, the average over time must be zero:

$$\begin{aligned} \frac{(1-\mu)TV_{in} + \mu T\left(V_{in} - V_{REF}\right)}{T} &= 0 \\ (1-\mu)\left(\frac{3}{2}\alpha\Delta V_{BE} - \frac{1}{2}V_{BE}\right) + \mu\left(\frac{1}{2}\alpha\Delta V_{BE} - \frac{3}{2}V_{BE}\right) &= 0 \end{aligned} \quad (4.40)$$

which results, when solved for μ, in

$$\mu = \frac{\frac{3}{2}\alpha\Delta V_{BE} - \frac{1}{2}V_{BE}}{V_{BE} + \alpha\Delta V_{BE}} \quad (4.41)$$

already derived in (4.36).

Every clock cycle thus, the integrated voltage is

$$3\alpha\Delta V_{BE} - V_{BE} - bs\left(2(V_{BE} + \alpha\Delta V_{BE})\right), \quad (4.42)$$

which means for $bs = 0$

$$3\alpha\Delta V_{BE} - V_{BE}, \quad (4.43)$$

while for $bs = 1$

$$3\alpha\Delta V_{BE} - V_{BE} - 2(V_{BE} + \alpha\Delta V_{BE}) = \alpha\Delta V_{BE} - 3V_{BE}. \quad (4.44)$$

This requires the implementation of the additional gain ± 3, used to multiply either $\alpha\Delta V_{BE}$ or V_{BE}, when $bs = 0$ or $bs = 1$, respectively. Also, it requires a different and longer $\Delta\Sigma$ cycle, since now, regardless of the value of bs, both $\alpha\Delta V_{BE}$ and V_{BE} need to be integrated every clock cycle.

4.3.4.4 V_{TEMP} implementation trade-off

Regardless of the implementation form, the result in (4.30) when compared with (4.43) for $bs = 0$ or the result in (4.31) when compared with (4.44) for $bs = 1$ states that, at best, the length of the original $\Delta\Sigma$ cycle is doubled, which nulls the conversion time reduction obtained with the new V_{TEMP} (according to what was presented in Table 4.1). This best situation is achieved if the gain 3 that multiplies either $\alpha\Delta V_{BE}$ or V_{BE} depending on bs is implemented at the expenses of a higher integration gain (which turns into a relatively larger front-end and integrators load). The worst scenario occurs if the gain 3 is implemented by multiple integration cycles, in which case the

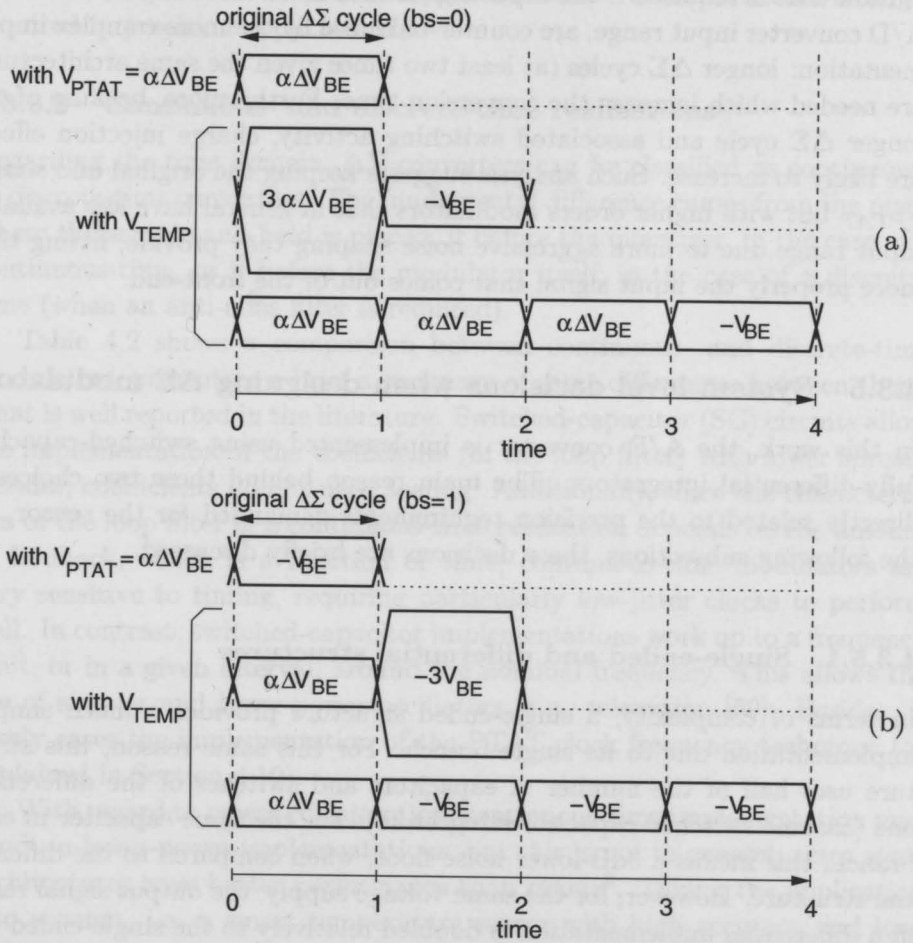


Figure 4.8: Original V_{PTAT} compared with two possible implementation of V_{TEMP}, either for (a) $bs = 0$ ($\alpha\Delta V_{BE}$ cycle) and (b) $bs = 1$ (V_{BE} cycle).

ΔΣ cycle is increased by a factor of 4 compared with the original V_{PTAT} cycle. Both situations are graphically presented in Fig. 4.8.

In conclusion, despite of using more of the A/D converter input range and keeping compatibility with the front-end architecture, the implementation of the voltage V_{TEMP} is more complex. Power savings, due to the reduced resolution that is required if the input signal uses more efficiently the available A/D converter input range, are counter-balanced by the more complex implementation: longer ΔΣ cycles (at least two times given the same architecture) are needed which increase the conversion time. Furthermore, because of the longer ΔΣ cycle and associated switching activity, charge injection effects are likely to increase. Such analysis suggests keeping the original and simple V_{PTAT} but with higher orders modulators that in general have less available input range due to more aggressive noise shaping they provide, fitting thus more properly the input signal that comes out of the front-end.

4.3.5 System level decisions when designing ΔΣ modulators

In this work, the A/D converter is implemented using switched-capacitor fully-differential integrators. The main reason behind these two choices is directly related to the precision requirements demanded for the sensor. In the following subsections, these decisions are briefly discussed.

4.3.5.1 Single-ended and differential structures

In terms of complexity, a single-ended structure provides a much simpler implementation due to its single branch. For this same reason, this structure uses half of the number of capacitors and switches of the differential one (assume switched capacitor integrators). For the same capacitor in each branch, this means a 3dB-lower noise floor, when compared to the differential structure. However, for the same voltage supply, the output signal range in a differential implementation is doubled relatively to the single-ended implementation. This means a maximum input signal 6dB-higher in the differential structure. Therefore, concerning the signal-to-noise ratio SNR, the differential structure has a 3dB-higher SNR (up to 4.5dB higher in some cases [50]).

For the same SNR, the size of the capacitors used in the differential structure can be halved, meaning for the same time constants, smaller switches sizes. This is fundamental in precision applications, since charge injection that comes from these switches - and introduces errors in the conversion - are proportional to their sizes.

Additionally, the differential structure helps to cancel the charge injected during switching, since this charge acts like a common-mode signal, as well to minimize other spurious effects like those that may come from the digital circuitry (coupled through the substrate) or from power supply.

Despite of the common-mode feedback required in the differential implementation, all other characteristics makes the differential structure strongly required in precision applications.

4.3.5.2 Continuous- and discrete-time realizations

Regarding the time domain, ΔΣ converters can be classified as continuous- or discrete-time converters. The fundamental difference comes from the place where the sample-and-hold is placed, if before the quantizer, in the case of a continuous-time, or if before the modulator itself, in the case of a discrete-time (when an anti-alias filter is required).

Table 4.2 shows a comparison between continuous- and discrete-time modulators, presenting a brief a summary of some differences between them, what is well reported in the literature. Switched-capacitor (SC) circuits allow the implementation of the coefficients (of the loop filter) with lower spread. Besides, coefficients do not need scaling. Additionally, since the characteristics of the loop filter in a continuous-time realization depends on the amount of feedback, which is a function of time, continuous-time modulators are very sensitive to timing, requiring particularly low-jitter clocks to perform well. In contrast, switched-capacitor implementations work up to a frequency limit, or in a given interval, around the nominal frequency. This allows the use of simpler and lower-power oscillators, e.g. relaxation [69]. Besides, it surely eases the implementation of the PTAT clock frequency technique (as explained in Section 4.10).

With regard to power consumption, continuous-time implementation may result in lower-power implementations, but this is not in general, since some architectures have better performance than others[8]. Taking the application into account, i.e. a smart temperature sensor with high accuracy and low-power operation, switched-capacitors integrators were understood having an overall superior performance. Consequently, they were chosen for this work.

4.3.5.3 Modulator Order

The modulator order is another important design parameter. The higher the modulator order is, the greater is the achievable resolution within a given

[8]Furthermore, if low-power operation is the most relevant specification, ΔΣ A/D converters may even not be the indicated architecture.

Table 4.2: Continuous- versus discrete-time modulators.

	Discrete-time	Continuous-time
Integrator type	SC	RC, gmC
Coefficient spread	low (capacitor ratio)	high (absolute RC)
Anti-alias filter	required	not required
Jitter sensitivity	lower	higher
Clock scaling	automatic	coefficients need scaling
Power consumption	low	lower

number of ΔΣ cycles. This suggests the use of of higher-orders modulators implementing the ΔΣ A/D converter. But the power due to extra circuitry in the modulator (and decimation filter) needs to be accounted for. This is analyzed as follows.

4.3.6 First, second and third-order ΔΣ modulators

In this section, first-, second- and third-order incremental ΔΣ modulators are analyzed and compared in terms of number of cycles needed for a given resolution. Power consumption is taken into account, using some of the results already presented in Subsection 4.3.2.

4.3.6.1 First-order ΔΣ modulator

A first-order ΔΣ modulator results if the loop filter $H(z)$ (as shown Fig. 4.3) given by

$$H(z) = \frac{z^{-1}}{1 - z^{-1}} = \frac{1}{z - 1}, \tag{4.45}$$

with a pole at DC, is used to realize the analog sampled-data integration. The equation describing the circuit is

$$Y(z) = [X(z) - Y(z)]\frac{z^{-1}}{1 - z^{-1}} + E(z) \tag{4.46}$$

which leads (if solved for $Y(z)$) to

$$Y(z) = z^{-1}X(z) + (1 - z^{-1})E(z). \tag{4.47}$$

By comparing (4.47) with (4.13), the Signal Transfer Function STF is found as z^{-1} while the Noise Transfer Function NTF as $(1 - z^{-1})$. The signal that appears at the output is just delayed by one clock period while the noise is passed through a difference which provides a first-order high-pass filtering.

Analyzing the NTF in the frequency domain (after z is replaced by $\exp^{j2\pi fT_S}$, $T_S = \frac{1}{f_S}$) results

$$\begin{aligned} NTF(f) &= 1 - \exp^{-j2\pi fT_S} \\ &= 2j \exp^{-j2\pi f\frac{T_S}{2}} \sin\left(2\pi f\frac{T_S}{2}\right), \end{aligned} \tag{4.48}$$

which shows the noise shaping a first-order ΔΣ modulator provides. While the quantization noise is amplified by 4, it is shaped by $\sin^2\left(2\pi f\frac{T_S}{2}\right)$ giving rise to a significant attenuation of the low-frequency components of the spectrum [66]. Restricting the bandwidth of the oversampled signal to f_B, a lower noise power will be present inside the signal band:

$$n_B^2 = \int_0^{f_B} \frac{\delta^2}{12}\frac{1}{f_S/2} 4\sin^2(\pi f T_S)\,\mathrm{d}f \tag{4.49}$$

$$\approx \frac{\delta^2}{12}\frac{1}{f_S/2} f_B^3 T_S^2 = e_{q,rms}^2 \frac{\pi^2}{3}\frac{1}{OSR^3} \tag{4.50}$$

since $\sin(\pi f T_S) \approx \pi f T_S$, for $OSR \gg 1$.

That is, oversampling with noise shaping reduces the in-band rms noise from ordinary quantization by the square root of OSR^3. The SNR in this case is given by

$$SNR = \frac{3}{2}\frac{3}{\pi^2} OSR^3 2^{2n} \tag{4.51}$$

or, in dB, as

$$SNR = 6.02n + 1.76 - 5.17 + 9.03\log_2(OSR)\ [\mathrm{dB}] \tag{4.52}$$

which means that for each doubling of the sampling frequency improves the ENOB by 1.5 bit.

Figure 4.9(a) and (b) plots the output spectrum of a first-order ΔΣ modulator when simulated with a sine wave and DC input, respectively. The noise shaping characteristic of 20dB/decade is consistent with a first order modulator (as well the harmonic content). While this result shows the noise-shaping effect, in order to get higher resolutions using a 1-bit quantizer[9], it is necessary to use a still fairly large OSR, since often the poorly shaped and large tones can fall in the signal band, which impairs the performance of the A/D converter [66]. Higher order modulators are thus often indicated.

The SNR in (4.52) expresses the performance of classical ΔΣ A/D converters under the assumption of busy input signals, which is not the case for

[9]Since noise shaping does not decrease the D/A converter linearity requirement, a two level quantizer, which is inherently linear, is assumed [72].

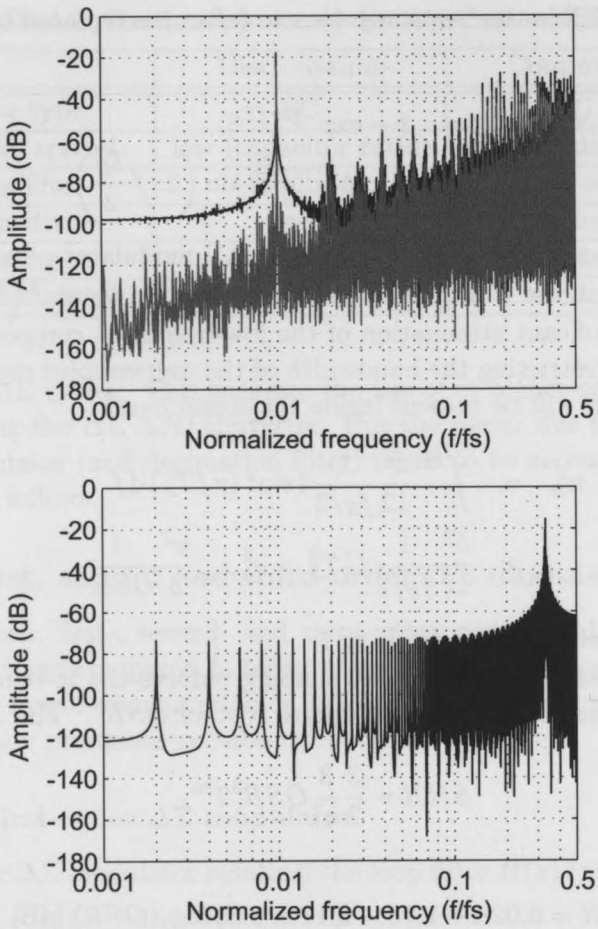


Figure 4.9: Simulated output spectrum of a first-order ΔΣ modulator (fft of 8192 points) with a) a sine input at $\frac{f}{f_S} = 0.01$ and amplitude of $\frac{1}{4}V_{REF}$ (used input range from 0.25% to 0.75% of V_{REF}) with and without Hanning windowing and b) a DC input of $\frac{1}{3}V_{REF}$.

a DC input. Besides, in incremental converters such performance is rather measured in number of bits, i.e. effective number of bits ENOB, despite of its relation with the SNR. The ENOB of a first-order ΔΣ modulator can be estimated using its time model as follows.

Resolution - Effective number of bits (ENOB) for DC input

The integrator input V_A, as shown in Fig. 4.10, is given by

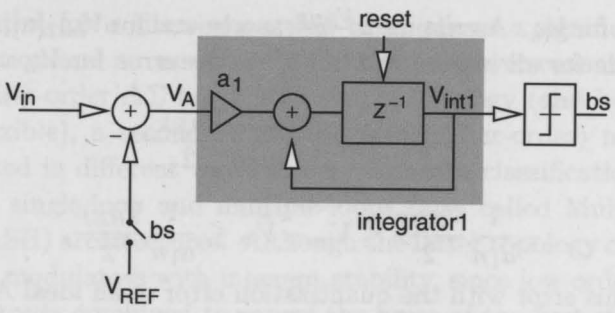


Figure 4.10: Discrete-time model of a first-order $\Delta\Sigma$ modulator.

$$V_A[n] = V_{in}[n] - bs[n]V_{REF} \tag{4.53}$$

while the output given by

$$v_{int1}[n] = v_{int1}[n-1] + a_1 V_A[n-1]. \tag{4.54}$$

After n cycles after reset ($v_{int1}[0] = 0$) the output of the integrator is given by

$$v_{int1}[n] = a_1 \sum_{k=0}^{n-1} (V_{in}[k] - bs[k]V_{REF}) \tag{4.55}$$

where $bs[k] = 0, 1$ is the comparator output, i.e. the bitstream value for the cycle k.

Assuming $V_{in}[k] = V_{in}$, i.e. constant input (DC signal), which is reasonable even if the input is varying slowly, the output of the integrator given in (4.55) can be rewritten as

$$v_{int1}[n] = a_1 n V_{in} - a_1 V_{REF} \sum_{k=0}^{n-1} bs[k]. \tag{4.56}$$

Rearranging this equation results in

$$\begin{aligned} \frac{v_{int1}[n]}{a_1 n} &= V_{in} - \frac{1}{n} V_{REF} \sum_{k=0}^{n-1} bs[k] \\ &= V_{in} - \hat{V}_{in} \end{aligned} \tag{4.57}$$

where

$$\hat{V}_{in} = \frac{1}{n} V_{REF} \sum_{k=0}^{n-1} bs[k] \tag{4.58}$$

is an estimate for V_{in}. Assuming $\pm\frac{V_{REF}}{2}$ as a bound for $V_{int1}[n]$, also assuming the loop stable for all inputs, the limits on the error for V_{in} can be written as

$$V_{in} - \hat{V}_{in} = \pm\frac{1}{a_1 n}\frac{V_{REF}}{2} \tag{4.59}$$

or

$$-\frac{1}{a_1 n}\frac{V_{REF}}{2} \leq V_{in} - \hat{V}_{in} \leq \frac{1}{a_1 n}\frac{V_{REF}}{2}. \tag{4.60}$$

Comparing this error with the quantization error of an ideal A/D converter, i.e. limited to $\pm\frac{V_{LSB}}{2}$, where $V_{LSB} = \frac{V_{REF}}{2^{nbit}}$, the following results (if $a_1 = 1$)

$$\frac{1}{n}\frac{V_{REF}}{2} = \frac{V_{LSB}}{2} = \frac{V_{REF}}{2.2^{nbit}}, \tag{4.61}$$

which allows to calculate the effective number of bits $ENOB_{1st,ideal}$ (n_{bit}) that ideally can be a achieved with a first-order modulator (assuming a matched decimation filter implemented according to (4.58):

$$ENOB_{1st,ideal} = \log_2(n). \tag{4.62}$$

This result shows that for each doubling of the number of cycles, 1 extra bit of resolution is obtained. It is worth comparing this result with the one given in (4.52), where 1.5 bit is obtained for the same frequency increase. In the case of an incremental A/D converter, the input is a DC signal, and because of this, very different from a busy signal. This means that the linear model used for the quantizer is no longer valid, since the quantization error in the case of a DC input can have a strong correlation with it [68]. This explains the difference. If a noise source is added right before the quantizer, a technique referred as dithering [63], a higher resolution in principle can be obtained.

Recalling to the result in (4.62), if a resolution of 15 bits is needed, $n = 2^{15}$ cycles are required, which makes the conversion time of a first-order A/D converter unacceptably long (the energy per conversion would be high regardless of a low power operation). The long conversion time can be circumvented by using a high operating frequency, but this increases bandwidth requirements for the integrators, e.g. for a sensor with a bandwidth of 10Hz, a clock frequency of 328kHz is needed. This suggests the use of higher order modulators [63][64]. A second-order (and a third-order) can reduce the number of cycles due to the more aggressive noise shaping they provide.

4.3.6.2 Second-order ΔΣ modulator

A second-order modulator provides the same resolution with much less clock cycles than a first-order modulator. Although not clear from (4.27), a higher-

order modulator, for a given conversion time, allows a significant clock frequency reduction and an associated reduction in power consumption [22].

While a first-order $\Delta\Sigma$ has a very simple topology (and because of that not really flexible), a second-order (and also higher-order) modulator can be implemented in different ways. A very common classification divides the modulator in single-loop and multiple-loops (also called Multi-stage Noise Shaping - MASH) architectures. Although the latter topology can implement higher-orders modulators with inherent stability, since low-order modulators have their outputs combined to cancel the noise of the first stage [60], they are not fully compatible with the charge balancing developed in Subsection 4.3.3[10]. The fact that the mismatch between the analog and digital parts can prevent such modulators of attaining the predicted high performance is another drawback. Because of that, single-loop architectures were considered more suitable for this work.

Among the single-loop topologies [74], it is worth to mention two in particular: Cascade of Integrators with Multiple Feedback - Input Feedforward (CIFB-IF) and Cascade of Integrators with Multiple Feedforward - Input Feedforward (CIFF-IF) [75]. They are depicted in Fig. 4.11(a) and Fig. 4.12, respectively. In both topologies, the output of the integrators are not dependent on the input signal[11], i.e. the integrators do not process the input signal at all. Because of this, the voltage swing at the integrators output is smaller and thus the distortion generated by the OPAMP. Additionally, because of the reduced output swing, low-power operation can be obtained. Besides both topologies are fully compatible with the charge balancing presented.

Perhaps the major drawback of both topologies compared to those without the distributed input-feedforward paths is the increased load the input (i.e. the sensor front-end) has to drive, particularly large for higher orders modulators [74]. In this case, the sampling capacitors have to be designed small, respecting the minimum sizes required by the application, so to avoid overloading the input. Luckily, the sizes of capacitors can be scaled down as the modulator order increases. Despite of both architectures having similar performances, the CIFB-IF structure was chosen to avoid the passive summation before the comparator the latter one requires.

[10]Because $\alpha\Delta V_{BE}$ is only input to the first stage in a MASH architecture, an explicit reference voltage V_{REF} is necessary in the subsequent stages.

[11]In the CIFB-IF architecture as shown, this is true only for the first integrator.

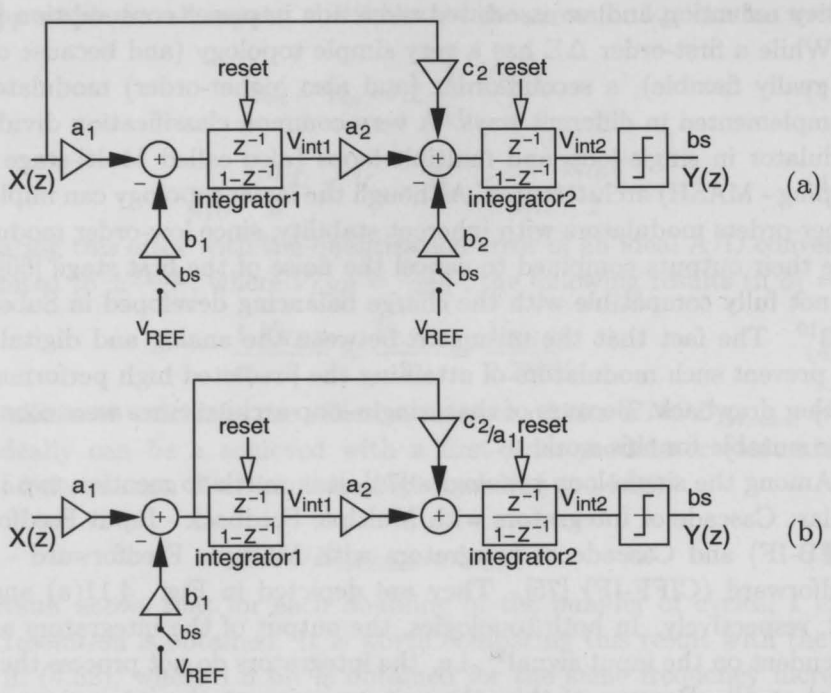


Figure 4.11: Second-order CIFB-IF $\Delta\Sigma$ modulator. a) Canonical form and b) equivalent form ($a_1 = b_1$ and $c_2 = b_2$).

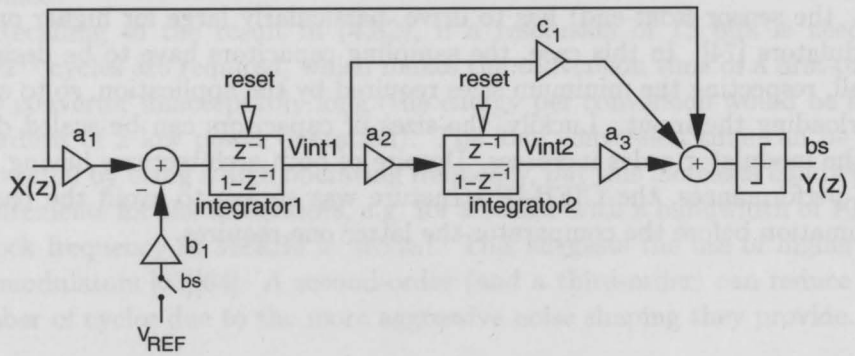


Figure 4.12: Second-order CIFF-IF $\Delta\Sigma$ modulator (canonical form).

Resolution for a DC input

The ENOB of a second-order modulator, as derived in the Appendix C, is given by

$$ENOB_{2nd,ideal} \approx 2\log_2(n) - 1 \qquad (n \gg 1). \tag{4.63}$$

This means that, for the same 15 bits of resolution, now only about 256 cycles are necessary. This number compared with the former 32768 cycles needed with a first-order ΔΣ modulator means is a huge shortening in the conversion time (about 99%), which is the main motivation to use a second-order modulator to save power. Given the same sensor bandwidth, e.g. 10Hz, an operating frequency less than 2.6kHz is now required.[12]

4.3.6.3 Third-order ΔΣ modulator

Since is the first integrator in a second-order modulator the circuit that more consumes (roughly 65% of the entire modulator's power consumption), the addition of a third integrator should add to a small fraction. On the other hand, the conversion time with a third-order modulator can be further reduced with respect to a second-order modulator for the same resolution due to the more aggressive noise shaping a third-order modulator can provide. Besides, the smaller available modulator's input range in a third-order modulator fits better the classical voltage V_{PTAT} and, in practice, a more efficient use of the available input dynamic (up to 60%) is observed. The need for a higher-order decimation filter is the main drawback of such approach. However, such filters are digitally implemented and often dissipate less than the modulator. It is this author's expectation that this choice provides indeed a better accuracy-power consumption trade-off. A brief study was included as follows as a proof of concept.

Resolution - ENOB

The ideal ENOB of a third-order modulator, as derived in Appendix C, achieved after n cycles is

$$ENOB_{3rd,ideal} \approx 3\log_2(n) - 2.58 \qquad (n \gg 1). \tag{4.64}$$

This means that, for 15 bits given as example, only about 58 cycles are necessary compared with the former 32768 and 256 cycles needed with a first- and second-order ΔΣ modulator, respectively. Figure 4.13 compares

[12] In practice, a larger number of cycles and thus frequency might be needed. These values though are still enormously smaller then those needed with a first-order modulator.

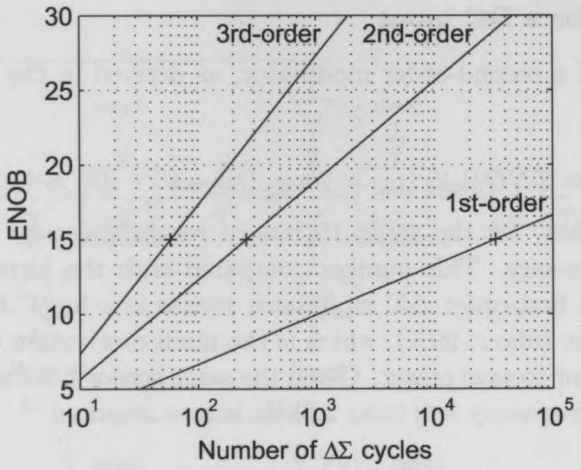


Figure 4.13: Achievable (ideal) ENOB versus number of ΔΣ cycles for a first-, second- and third-order ΔΣ modulator.

the achievable (ideal) ENOB versus the number of ΔΣ cycles for a first-, second- and third-order ΔΣ modulator.

The main drawbacks of higher orders modulators are the power consumption of the additional integrators and the more complex implementation of the decimation filter, since the filter order follows the modulator order [63]. However, the estimated power consumption of a third-order modulator is about only 15% larger than that of a second-order modulator, since it is the first integrator the most power consuming circuit, while the number of cycles required decreases by about 77% (for a resolution of 15 bits). This clearly shows the advantage of a third-order modulator when compared with a second-order one, probably enough to compensate the more complex decimation filter implementation.

Design of a third-order ΔΣ Modulator

In this section, some important aspects related to a third-order ΔΣ modulator, as shown in Fig. 4.14, whose system level design is detailed in Appendix D, are presented. Because the input signal that comes from the front-end uses only about 30% of the available A/D converter input range, a third-order ΔΣ modulator fits better the input signal, where an input range normally less than 75% of V_{REF}, instead of 90% in the case of a second-order, is available [63]. Although this modulator was not integrated in any circuit and no measurements exist, it is presented here as a concept to show that the same

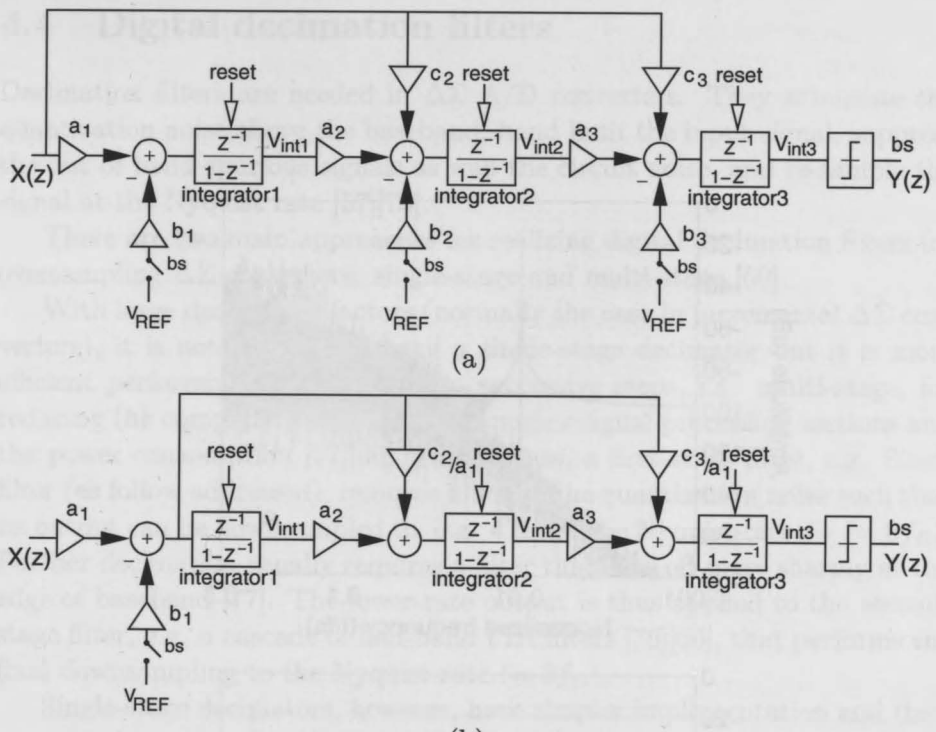


Figure 4.14: Third-order CIFB-IF ΔΣ modulator. (a) Canonical form and (b) equivalent form ($a_1 = b_1$, $c_2 = b_2$ and $c_3 = b_3$).

charge balancing and topology used with a second-order modulator is also feasible with a third-order modulator (also with higher-orders modulators having the same topology, i.e. CIFB-IF).

The simulation result (bitstream spectrum) is presented in Fig. 4.15. This shows that a third order modulator, whose input range fits more efficiently the input signal (the voltage V_{PTAT}), can successfully implement the classical charge balancing as described. Taking into account also the lower number of cycles for a given resolution, this makes a third-order modulator even more affordable to implement low-power temperature sensors. However, the effect of the reduced number of cycles on DEM and thus on the sensor accuracy needs to be analyzed. This is later commented in Section 4.6.2.

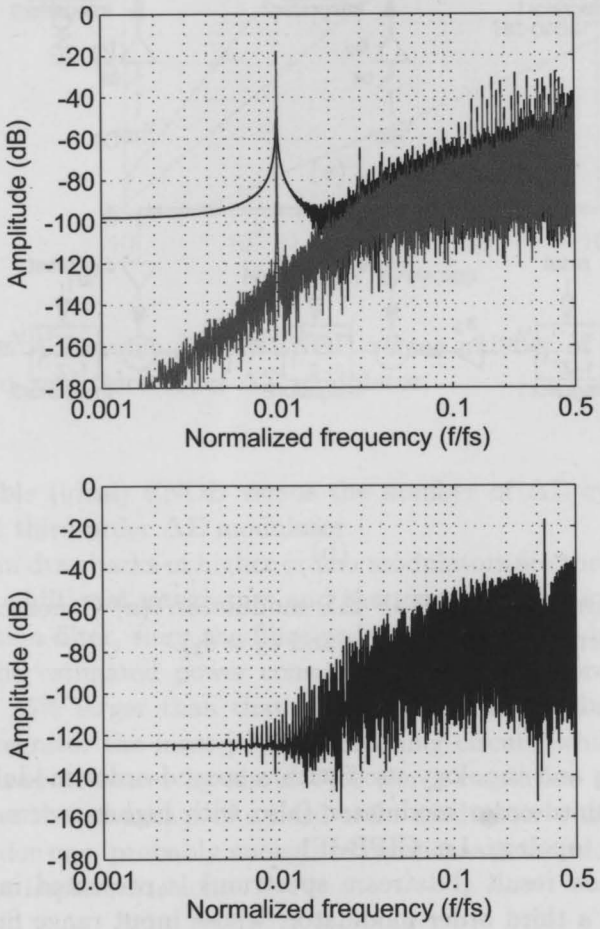


Figure 4.15: Simulated output spectrum of a third-order $\Delta\Sigma$ modulator (fft of 2048 points) with a) a sine input at $\frac{f}{f_S} = 0.01$ and amplitude of $0.375V_{REF}$ (used input range from 0.35% to 0.75% of V_{REF}) (without and with windowing - Hanning) and b) a DC input of $\frac{1}{3}V_{REF}$.

4.4 Digital decimation filters

Decimation filters are needed in $\Delta\Sigma$ A/D converters. They attenuate the quantization noise above the baseband, band limit the input signal, suppress the out of band spurious signals as well the circuit noise, and re-sample the signal at the Nyquist rate [57][76].

There are two main approaches for realizing digital decimation filters for oversampling $\Delta\Sigma$ converters: single-stage and multi-stage [60].

With large decimation factors (normally the case in incremental $\Delta\Sigma$ converters), it is not convenient using a single-stage decimator but it is more efficient performing the operation in successive steps, i.e. multi-stage, for reducing the complexity of the high-frequency signal processing sections and the power consumption [77][66]. In this case, a first-stage filter, e.g. Sinc^L filter (as follow addressed), removes much of the quantization noise such that its output can be downsampled to, e.g. 4 times the Nyquist rate f_N $(= 8f_B)$. Further decimation usually requires a filter that cuts off more sharply at the edge of baseband [77]. The lower-rate output is thus applied to the second-stage filter, e.g. a cascade of half-band FIR filters [76][60], that performs the final downsampling to the Nyquist rate $(= 2f_B)$.

Single-stage decimators, however, have simpler implementation and their use with incremental $\Delta\Sigma$ converters is also reported [63][22]. In this case, the decimator performs the downsampling to the Nyquist rate f_N in just one step. The decimation filter is either implemented by a direct calculation using cascade of integrators (W_L) or Sinc^L filters.

Although within the scope of this work, a power consumption analysis of decimation filters (often close to $\frac{1}{3}$ of the total A/D converter power consumption) needs a specific work and thus was skipped here. Considerations on the chosen single-stage Sinc^L filters, however, are presented with some design considerations.

4.4.1 Direct calculation of the digital output

The conceptually simplest filter to produce the digital output of the incremental A/D converter is a cascade of integrators [63] implementing the formula given in (4.58) for a first order modulator or in (C.10) and (C.27) for a second- and third-order CIFB-IF topologies modulators [13] (refer to Appendix C). While in a first-order modulator, this filter can be easily implemented as a counter, digital integrators are needed for higher-orders filters. Fig. 4.16(a) shows the normalized impulse response of three cascade-of-integrators fil-

[13] Because the formula is derived for a particular modulator topology, such filters are sometimes referred as matched filters.

ters, namely W_1, W_2 and W_3, derived for the topologies discussed, e.g. first, second- and third-order CIFB-IF, respectively, while Fig. 4.16(b) shows their simulated frequency response (magnitude).

Moreover, for higher-order incremental converters, because of the output of the modulator which is not simply averaged, but repeatedly accumulated by means of several integrators, a periodic noise (e.g. line frequency and interference due to DEM and chopping) cannot be suppressed using a direct-implementation filter since the weighting process varies asymmetrically from sample to sample. Other drawbacks are the highest weight they put on the first samples and the poor frequency response. On the other hand Sinc^L filters can provide periodic noise suppression, since they have a symmetrical impulse response, with the the highest weights in the middle. Furthermore, such filters can preserve the rising spectral shape of the noise from the ΔΣ modulator, keeping most of its power at high frequencies [77].

4.4.2 Filtering using Sinc^L filters

In a first-order incremental ΔΣ A/D converter, the decimator is a simple counter that is reset before each conversion. The counter averages out the data from the noise shaper over the bitstream and converts (down samples) an N-bit bitstream input to a single word (decimated result). In other words, the counter implements a common technique for smoothing a noisy data sequence, i.e. sequence averaging over M adjacent data samples:

$$y[n] = \frac{1}{M} \sum_{i=0}^{M-1} x[n-i] \tag{4.65}$$

where M is the decimation factor ($M = N$ in case of a first-order incremental ΔΣ A/D converter). Such filter is known as Sinc filter[14].

For higher order ΔΣ A/D converters though, this simple and straightforward Sinc filter is not efficient since it does not provide sufficient noise suppression to prevent a significant increase in the baseband noise floor [76]. Higher order Sinc filters, named Sinc^L, are required. A Sinc^L filter is indeed a cascade of L Sinc filters. The corresponding impulse response in the z-domain $H(z)$ of an L^{th}-order modulator with a noise shaping function $(1 - z^{-1})^L$ and its frequency response (magnitude) obtained by evaluating $H(z)$ for $z = \exp^{j2\pi fT_S}$, as shown in 4.16(a) and (b) for $L = 1$ to 3, are expressed as

[14]The Sinc filter, also known as rectangular window, is not coincidentally the same matched filter W_1.

$$H(z) = \left(\frac{1}{M}\sum_{i=0}^{M-1} z^{-i}\right)^{L} = \frac{1}{M^{L}}\left(\frac{1-z^{-M}}{1-z^{-1}}\right)^{L} \tag{4.66}$$

and

$$|H(f)| = \frac{1}{M^{L}}\left(\frac{\sin(\pi f M T_S)}{\sin(\pi f T_S)}\right)^{L}, \tag{4.67}$$

respectively. The filter has $(M-1)$ zeros[15] at frequencies $k\frac{f_S}{M}$, for $k = 1, 2, ..., M-1$ with the first notch f_1 at $\frac{f_S}{M}$, hence considered the edge of the stopband of the filter.

It is worth to mention that in incremental $\Delta\Sigma$ A/D converters, the order of the filter needs to be at least as high as that of the modulator, but preferably higher by 1 [77][65][63]. This is explained by the fact that the slope of the attenuation for the low-pass (decimation) filter should be greater than the rising quantization noise, so that the resulting noise falls off at a relatively low frequency [60]. Nonetheless, observed measurement results with the decimation filter having the same order of the modulator presented comparable performance.

Additionally, since incremental converter operates in a transient mode, the transients of both analog and digital component blocks must settle sufficiently to achieve the specified accuracy [63]. For the digital filter, this means that all memory elements must be filled with valid data, i.e. the filter must be operated for at least $LM - (L-1) \approx LM$ clock cycles, in order to full fill all registers. Because there is a minimum number of clock cycles N so to achieve the necessary converter resolution, the following results:

$$N \geq LM \tag{4.68}$$

which sets the decimation factor, i.e. $M = N/L$. In other words, once the required number of cycles N is determined according to resolution needs, the needed decimation factor M is set (given the modulator order L).

4.4.3 Non-linear filtering

As discussed in Chapter 3, second-order non linearities can be compensated by using a PTAT current to bias the PNP transistors and by properly choosing the bias value so the reference voltage V_{REF} is slightly temperature dependent. However, a third-order remaining non linearity (which is responsible for a temperature error of about $\pm 0.15°C$) appears once the second-order non-linearity was removed. Therefore, a third-order curvature correction is

[15] The zero at DC ($z = 1$) for $k = 0$ is canceled by a pole at the same location.

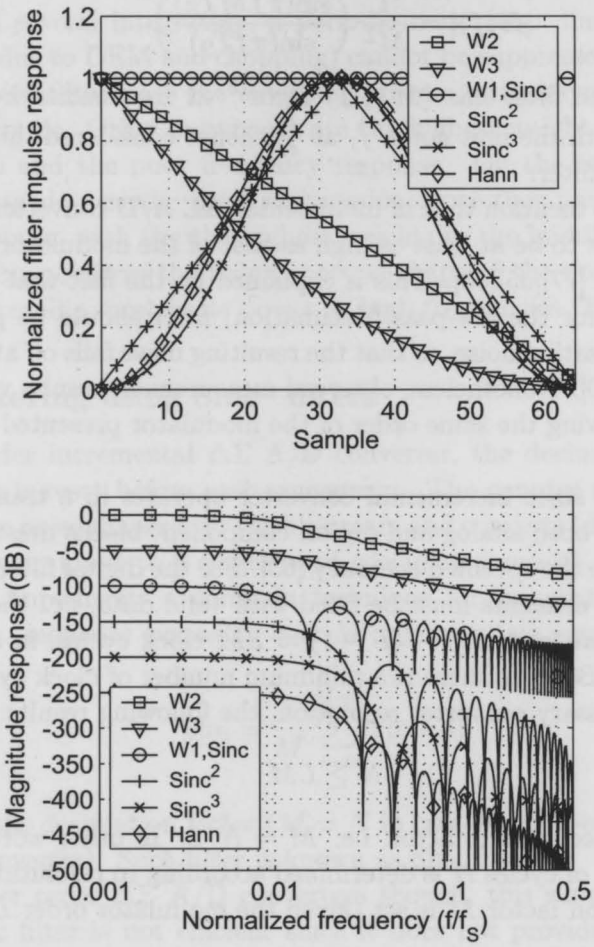


Figure 4.16: (a) Normalized filter impulse response of W_L and Sinc^L filters for $L = 1$, 2 and 3 ($N = 64$) and (b) their simulated frequency response (magnitude). Responses were vertically shifted for clarity.

needed to achieve an inaccuracy below 0.1°C over the entire military range. Such nonlinear correction can be implemented, for instance, as a look-up table [22], since the third-order non linearity has a very systematic and predictable behavior, according to many observations done in many sensors. It is out of scope to further discuss this topic in this work.

4.5 Modulator error sources

The A/D converter, in particular the modulator, has several error sources that, if not properly handled, will cause conversion errors impairing the sensor performance. This is regardless of how accurate the front-end is. Therefore, it is important to find how each error affects the modulator operation, looking for techniques to keep the errors bounded according to what the application requires.

The methodology here follows the same of that presented in Section 3.3 of the previous chapter. In this case, however, instead of focusing on the minimum PNP bias currents, it focus on the minimum modulator's sampling capacitors given accuracy requirements to reduce front-end's and modulator's load, allowing thus a low power operation. The same requirements as described in Chapter 3 applies unaltered.

4.5.1 Offset error

Input (e.g. V_{BE} or ΔV_{BE}) sampling is a fundamental operation prior to the A/D conversion. Therefore, it is important to perform it accurately. One typical sampling error comes with the OPAMP offset, often part of the circuitry used in the sampling operation. Because it adds directly the to input voltage, the same limit given in (3.15) applies, i.e.

$$|V_{OS}| < \frac{3\text{mV}}{\text{°C}} \frac{1}{\alpha} \varepsilon_T. \tag{4.69}$$

This means that for $\varepsilon_T = \pm 0.01$°C and $\alpha = 10$, a maximum error (due to offset) equal to $\pm 3\mu$V is allowed. This value is much smaller than typical offset voltages observed in CMOS OPAMPs [78], which suggests the use of techniques for offset cancellation. Because of the very low offset required, even the residual offset after offset cancellation might be of concern.

4.5.2 Gain error due to capacitor mismatch

A gain error, on the other hand, is due to capacitor mismatch used to implement the integration gains of the modulator, e.g. in particular α used to

implement the relative gain between ΔV_{BE} and V_{BE}. According to (3.16), in order to prevent an error greater than ε_T, the relative capacitor mismatch should be smaller than

$$\left|\frac{\alpha - \alpha_{ideal}}{\alpha_{ideal}}\right| < \frac{2}{3}\%\frac{1}{^\circ\mathrm{C}}\varepsilon_T \tag{4.70}$$

which means that, for $\varepsilon_T = \pm 0.01^\circ$C, a maximum α mismatch of 0.0067% is allowed. According to [28], a $20\mu\mathrm{m} \times 20\mu\mathrm{m}$ (≈ 0.3pF) capacitor has a relative mismatch of 0.11% which is far above the required value. This indicates that such capacitor match cannot be achieved by precise layout techniques only.

Evidently, the relative error due to mismatch can be reduced by increasing the capacitor sizes, but this often happens at the expenses of increased front-end and modulator load, and in consequence, power consumption. DEM technique is thus indicated to reduce the capacitor mismatch, avoiding gain error with no or small impact in area and power.

4.5.3 Error due to non-complete settling time

Switched capacitor integrators, while insensitive to coefficient spread and clock jitter, need time for proper settling. Otherwise, an error due to non-complete settling occurs, which adds to an error in the temperature reading. The minimum time t_S required for settling is given by

$$t_S \geq N_\tau . \tau$$

where N_τ

$$N_\tau = \ln(\frac{1}{\varepsilon}). \tag{4.71}$$

is the number of time constants necessary to settle within a given relative error ε[16] and τ is the time constant of the circuit given by[17]

$$\tau = \left(\frac{V_T}{I_{bias}} + R_{on}\right)\frac{\alpha C_S}{N_\alpha} \tag{4.72}$$

$$\approx \left(\frac{V_T}{I_{bias}}\right)\frac{\alpha C_S}{N_\alpha} \qquad \left(\frac{V_T}{I_{bias}} \gg R_{on}\right) \tag{4.73}$$

[16] Alternatively, the number of time constants necessary to settle to a B bits is given by $N_\tau = ln(2^B)$. For $B \approx 15$ bits, $N_\tau \approx 10$.

[17] In order to integrate ΔV_{BE} with a gain α times greater than the integration gain used for V_{BE}, a sampling capacitor α times greater is used. However, it is possible to use a smaller capacitor, i.e. $\alpha C_S/N_\alpha$, while performing multiple (N_α) integration cycles within a $\Delta\Sigma$ cycle.

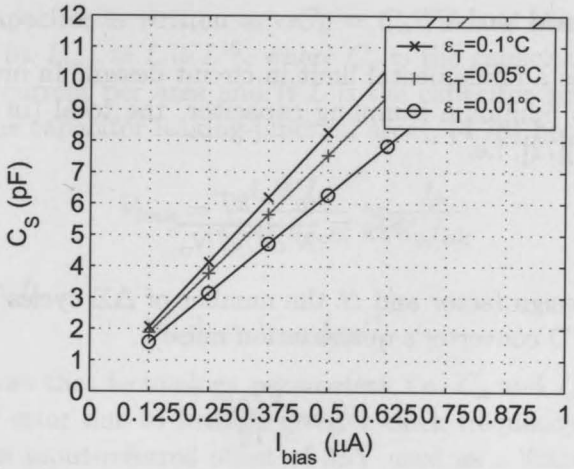


Figure 4.17: Maximum sampling capacitor C_S as a function of the error contribution ε_T and I_{bias}(f_{CLK} = 10kHz, $\alpha = 16$ and $T = 300$K).

where R_{on} is the switch on-resistance, V_T is the thermal voltage and I_{bias} is the smallest current used to bias the two PNP, when the front-end is generating the voltage ΔV_{BE}.

Assuming that the time available for settling is equal or smaller than a half-period of the system clock or N_α shorter if multiple cycles are used to integrate ΔV_{BE}, i.e. $\frac{T_{clk}}{2N_\alpha} = \frac{1}{2f_{clk}N_\alpha}$, and that errors due to incomplete settling have the same effect of α-mismatch errors (so the same limit can be used here), the following condition

$$\alpha C_S \leq \frac{I_{bias}}{2f_{clk}V_T \ln(\frac{1}{\varepsilon})} \tag{4.74}$$

results, which sets, given an operation frequency and an error limit (regardless of N_α), a maximum sampling capacitor for a given PNP bias current I_{bias} or, alternatively, a minimum PNP bias current given a front-end load capacitance (modulator's sampling capacitors). For a clock frequency f_{clk} = 10kHz, $\varepsilon = 0.0067\%$, (± 0.01°C) a capacitor $\alpha C_S \leq$ 50pF (at room temperature) is required if I_{bias} = 250nA. For $\alpha = 16$, this results in a $C_S \approx$ 3pF. Figure 4.17 presents some results for C_S obtained from (4.74) for some values of I_{bias} and temperature errors.

4.5.4 Thermal noise

Thermal noise is a fundamental limit in circuit design. In order to estimate the integrator's minimum sampling capacitor, the total (in band) thermal noise power [50][72], i.e.

$$S_{Th}^2 = \frac{m}{N}\frac{kT}{\alpha C_S/N_\alpha}, \tag{4.75}$$

where m is a design factor and N the number of $\Delta\Sigma$ cycles per conversion, is set to the A/D converter's quantization noise[18],

$$S_Q^2 = \frac{V_{LSB}^2}{12}. \tag{4.76}$$

For example, a resolution of 15 bits (LSB$\approx$ 0.02°C) gives a quantization noise of about $10.5\mu V_{rms}(V_{REF} = 1.2V)$. Assuming $m = 4$ and $N = 300$, a capacitor $\alpha C_S \geq 0.5$pF (for $N_\alpha = 1$) is required. For such capacitor, according to (4.74), settling time constraints imply a PNP bias current of a few nA ($\approx$ 2.5nA), which is far smaller than, for instance, the values found in Chapter 3 taking other errors into account.

4.5.5 Error due to leakage current

Leakage is meanly due to the reverse-biased p-n junctions, found in the drain- or source-bulk junctions of switches as well as in the diffusion-bulk of MOS capacitors. The leakage current rapidly increases with temperature (it approximately doubles for every temperature increase of 11°C in silicon [60]) and is responsible for errors when allows the voltage stored in the integration capacitors to leak. The charge lost within one clock cycle is

$$\Delta Q_{leak} = I_{leak}T_{clk} = \frac{I_{leak}}{f_{clk}} \tag{4.77}$$

where I_{leak} is the leakage current and f_{clk} the clock frequency (indeed N_α times f_{clk} if N_α integration cycles are used).

The input-referred offset voltage error due to leakage V_{leak} is found dividing ΔQ_{leak} by the sampling capacitance $\alpha C_S/N_\alpha$ (which is used to integrate ΔV_{BE}), i.e.

$$V_{leak} = \frac{\Delta Q_{leak}}{\alpha C_S/N_\alpha} = \frac{I_{leak}}{\alpha C_S f_{clk}}. \tag{4.78}$$

[18] A common approach is to design the $\Delta\Sigma$ A/D converter with its thermal noise slightly higher than the quantization noise, so its overall noise performance is mainly determined by the first one.

If the MOS capacitor is written as $\alpha C_S = C_o WL$ and the leakage current approximated by $I_{leak} \approx I_o WL$[19], where C_o is the capacitance per area, I_o is the leakage current per area and WL is the capacitor active area (which sets roughly the capacitor leaking-junction area), (4.78) becomes

$$V_{leak} = \frac{I_o WL}{C_o WL f_{clk}} = \frac{I_o}{C_o f_{clk}} \tag{4.79}$$

or, if solved for I_o,

$$I_o = V_{leak} f_{clk} C_o. \tag{4.80}$$

This result shows that technology parameters, i.e. C_o and I_o, largely set the the amount of error due to leakage given a clock frequency. Alternatively, for a maximum input-referred offset of 3μV used as a limit for V_{leak} (refer to (4.69)), a clock frequency of $f_{clk} = 10$kHz and a capacitance per area $C_o = 0.75\text{fF}/\mu\text{m}^2$ [28]), a maximum leakage current per area of $22.5\text{aA}/\mu\text{m}^2$ is estimated. This value though is much smaller than typical leakage currents per junction area I_o (about $0.15\text{pA}/\mu\text{m}^2$ at 140°C [28]) found in the present technology and therefore gives the magnitude of the leakage problem at high temperatures.

Obviously, this is a worst-case scenario since not necessarily the whole leakage current is translated to ΔQ_{leak} (e.g. part of it may be subtracted from the front-end bias current or be provided by the OTA). Additionally, leakage is predominantly a common-mode effect in a differential architecture, and mainly the mismatch between the leakage currents in both branches is of concern. Furthermore, together with a system level chopping, discussed in Section 4.8, a leakage current per area two orders of magnitude higher, e.g. $2.25\text{fA}/\mu\text{m}^2$, is more realistic. Nonetheless, such current is still smaller than I_o, suggesting that leakage is a major limiting factor of performance for temperatures above 125°C. And because the leakage current increases as the sampling capacitor increases, no improvement is observed if capacitor sizes are increased.

Finally, it is worth to mention that the leakage effect can be reduced by reducing the leaking time. This is achieved by increasing the system clock frequency, reducing ΔQ_{leak} as clear from (4.77). Nevertheless, restrictions due to other requirements, e.g. settling time, apply, limiting an arbitrary clock frequency increase.

[19] In the present technology, a minimum switch has a drain/source area equal to $5.5\mu\text{m}^2$, while, for instance, a 1pF MOS capacitor has an area greater than $1000\mu\text{m}^2$ (for a capacitance per area of $C_o = 0.75 f\text{F}/\mu\text{m}^2$ [28]). In this sense, the leakage current in the switches are neglected here. A different situation comes if double poly capacitors are used, when the leakage through the switches may dominate.

4.5.6 Error due to charge injection

Charge injection results from the switching mechanism. When a MOS switch opens, part of the charge that is accumulated under the gate or transistor channel ($Q_{channel}$) flows out through the transistor drain and source, modifying the charge stored in capacitors connected to the switch, often altering the sampled voltage. This is shown in Fig. 4.18. Clock feed-through [76] also contributes to the total charge injected, but it was not considered here.

Because the existence of MOS switches and sampling capacitors which are typical of the so called switched-capacitor circuits, errors due to charge injection often appear. In general, such errors can be minimized by using large capacitors and minimum switches, but requirements set by conversion or settling time can severely limit the use of large capacitors and switches with high-ohmic R_{on} resistances. More interesting is the fact that, once errors due to offset were mitigated, charge injection errors are the remaining problem [43], despite of differential architectures, minimum-size switches and other advanced techniques [22].

The total charge stored under the gate of a MOS switch is given by

$$Q_{channel} = C(V_{GS} - V_{Th}) \tag{4.81}$$

where C is mainly the gate capacitance given by

$$C = WLC_{ox} \tag{4.82}$$

where WL is the gate (or channel) area, C_{ox} is the oxide capacitance per unit area and $(V_{GS} - V_{Th})$ is the effective voltage over the gate. For $C_{ox} = 2\text{fF}/\mu\text{m}^2$ [28] and a minimum switch ($WL = 1.0\mu\text{m} \times 0.7\mu\text{m}$), the gate capacitance is $C = 1.4\text{fF}$. This results in a channel charge $Q_{channel} \approx 2.5\text{fC}$ for a $V_{GS} - V_{Th} = 1.8\text{V}$[20].

This result means that, in order to prevent a voltage variation below the input-offset voltage V_{OS}, used as an error limit due to charge injection in the sampled ΔV_{BE}, i.e. $V_{error} = 3\mu\text{V}$, a minimum sampling capacitor

$$\frac{\alpha C_{S,min}}{N_\alpha} = \frac{Q_{channel}}{V_{error}} = \frac{2.5f}{3\mu} \approx 800\text{pF} \tag{4.83}$$

or $C_S \approx 200\text{pF}$ (for $\alpha = 8$ and $N_\alpha = 2$) is needed.

Techniques like delayed clocks (important to prevent the charge injection dependency on the input signal) also help to alleviate the impact of charge injection, preventing the charge of switches controlled by the delayed clocks

[20] With a supply voltage higher than 2.5V, charge injection can indeed present higher values.

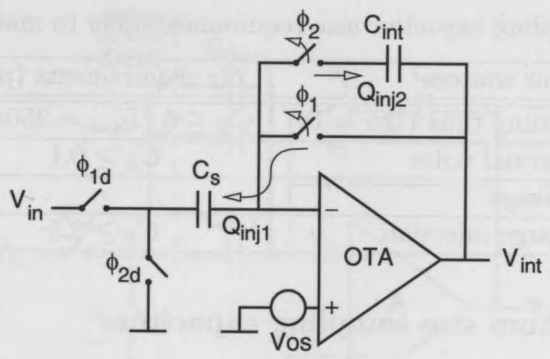


Figure 4.18: Integrator circuit schematic showing the charge injection due to switch opening. Phases ϕ_1 and ϕ_2 are non-overlapping, while phases ϕ_{1d} and ϕ_{2d} are both opening-delayed versions of ϕ_1 and ϕ_2, respectively.

of migrating to the sampling capacitors, but with no impact over the charge injected by the remaining switches (refer to Fig. 4.18).

Integrators with differential architecture, on the other hand, can significantly reduce the effect of charge injection by a factor 10, since in this architecture charge injection as leakage acts as a common-mode signal. In this case a minimum capacitor $C_{S,min} \approx 20$pF is expected. Advanced techniques, e.g. chopping of switches, are necessary though to further reduce the error introduced by charge injection. If one order of magnitude reduction is assumed once more, a minimum sampling capacitor $C_S \approx 2$pF is found. Two orders of magnitude however are indeed too optimistic. A more realistic factor, not greater than 80, is expected. For such factor, a capacitor $C_S \geq 2.5$pF is needed.

4.6 The sampling capacitor

The power consumed in a switched-capacitor integrator is generally proportional to its load. This load, or part of it is also load to the front-end. So, in order to minimize the power consumption in the integrator and front-end, capacitor sizes need to have minimum values for which the required converter resolution and bandwidth are still maintained. This approach often leads to a modulator design for which the performance is limited primarily by $\frac{KT}{C}$ noise in the first integrator [56]. In this work, however, due to particular accuracy requirements (often translated to matching needs), other error sources like charge injection and leakage are more prone to limit the sensor performance than thermal noise.

Table 4.3: Sampling capacitor size requirements due to main error sources.

Error sources	C_S requirements (pF)
Settling time (@$\alpha = 16$)	$C_S \leq 3$ ($I_{bias} = 250$nA)
Thermal noise	$C_S \geq 0.1$
Leakage	-
Charge injection	$C_S \geq 2.5$

4.6.1 Minimum size sampling capacitor

Given a conversion time and a temperature error contribution, the bias current of the PNP transistors in the front-end and the modulator sampling capacitors are immediately related through settling time requirements. Nonetheless, the design of the A/D converter follows other requirements, set by thermal noise, leakage and charge injection, as previously addressed, that turns into additional requirements to C_S. Based on all these requirements, which are summarized in Table 4.3, a best trade-off sampling capacitor (for the first integrator) was determined: 2.5pF. Smaller sampling capacitors, e.g. close to 1pF, were used in some designs but with visible lack of performance (refer to Chapter 5 for measurement results).

4.6.2 Designing an accurate capacitor - DEM

Dynamic element matching (DEM) as explained in Chapter 3 is a well-known technique [36][37] used to improve the effective accuracy of a system. It is particularly helpful when there are several nearly equal circuit elements that can be dynamically interchanged over time (here the sampling capacitors instead of the bias current sources), so mismatching errors are averaged out. The effective accuracy is often some orders of magnitude higher than the relative accuracy of such elements themselves [15], so a matching requirement of 0.0067% can be obtained from typical levels of capacitor matching, e.g 0.1% [28][35].

Figure 4.19 shows how the sampling capacitors are dynamically interchanged. Every V_{BE} cycle, a capacitor $C_{S,j}$

$$C_{S,j} = C_S + \Delta C_{S,j} = C_S\,(1 + \delta_j) \qquad (1 \leq j \leq \alpha) \tag{4.84}$$

where $\delta_j = \frac{\Delta C_{Sj}}{C_S}$ is the relative mismatch error of the j^{th} capacitor with respect to the average capacitance C_S, is used as the sampling capacitor, while the remaining capacitors are switched out of the circuit. The capacitor selection is digitally controlled and occurs in between the non-overlapping phases ϕ_1 and ϕ_2, avoiding disturbing the integration operation. In a ΔV_{BE}

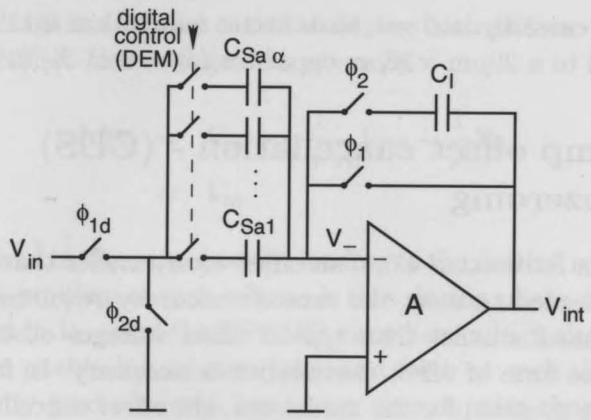


Figure 4.19: First integrator schematic showing α capacitors used to set an accurate 1 to α capacitor ratio (relative integration gain). Switches (digitally controlled) are used to dynamically interchange the capacitors, so to implement DEM. Clock phases ϕ_1 and ϕ_2 are non-overlapping.

cycle, all capacitors are selected, resulting in a parallel capacitor αC_S , which is α times the average capacitor C_S.

Because

$$\sum_{j=1}^{\alpha} \delta_j = 0, \tag{4.85}$$

mismatching errors are completely averaged out every α $\Delta\Sigma$ cycles. However, according to [22], if the number of V_{BE} cycles is not a multiple of α, there will be a remaining gain error after N cycles, bounded as

$$\left| \frac{\alpha - \alpha_{ideal}}{\alpha_{ideal}} \right| \leq \frac{\sqrt{\alpha}}{N} \delta_{max} \tag{4.86}$$

where δ_{max} is the worst- (or initial) case mismatch, i.e. all $|\delta_j| < \delta_{max}$. If, for example, $\alpha = 8$ and $\delta_{max} = 0.11\%$, at least $3N = 130$ $\Delta\Sigma$ cycles per conversion are needed to obtain a mismatching error below 0.0072% (α-mismatch requirement at $T = -55°\text{C}$, when a V_{BE} cycle happens approximately once every 3 $\Delta\Sigma$ cycles since $\mu = \frac{1}{3}$).

Interesting to mention that matching requirements, setting a minimum number of $\Delta\Sigma$ cycles, prevent in practice the use of modulators with orders higher than three. For instance, while a third-order modulator needs a few tens of cycles (58) to achieve a resolution of 15 bits, 130 cycles are still needed to keep mismatching errors bounded to the level given in (4.86). A third order though is still acceptable since a sampling capacitor in the order

of, e.g. 2.5pF, carefully laid out, have better match than 0.11% due to larger area compared to a 20μm $\times$20μm capacitor ($\delta_o = 0.11$ %) [28].

4.7 Opamp offset cancellation - (CDS) Autozeroing

As discussed in Subsection 4.5.1, an offset error smaller than 3μV (as given in (4.69)) is needed to meet the sensor's accuracy requirements. Because this value is much smaller than typical offset voltages observed in CMOS OPAMPs, some form of offset cancellation is necessary. In face of the total current budget specified for the modulator, the offset cancellation technique has to be simple, avoiding additional nulling amplifiers (chopper-stabilization [24][78]) that increase complexity and often contribute to the total amplifier's current consumption.

Autozeroing [24][43][78], instead of chopping – used for offset cancellation of the front-end bias OTA (refer to Subsection 3.6.3 of Chapter 3), is chosen for the modulator. This is because of its implementation which is simple, not requiring another amplifier. But also because it employs the same sampling capacitor, used to sample the A/D converter input signal, to also sample and store the OPAMP's offset V_{OS} (i.e. input offset sampling).

The open-loop autozeroing offset cancellation principle is not well suited to high-gain amplifiers, since a small offset voltage can cause the amplifier to saturate [43]. A closed-loop configuration (e.g. with the amplifier in unity-gain) is then usually preferred to sense the amplifier's offset. A particular case of autozeroing is the well known Correlated Double Sampling (CDS), which is shown in Fig. 4.20. During the sampling phase $\phi 1$, the amplifier is in unity-gain feedback. Assuming that the open-loop gain A of the amplifier is much larger then one, i.e. $A \gg 1$, the sampled voltage V_S across the sampling (and offset storage) capacitor C_S is[21]

$$V_{S,\phi 1} = V_{in} - \frac{A}{1+A} V_{OS} \tag{4.87}$$

$$\approx V_{in} - V_{OS} \qquad (A \gg 1). \tag{4.88}$$

During the integration phase $\phi 2$, this voltage, added of V_{OS} is available for amplification. The capacitor C_{int} is switched in the feedback path and C_S is discharged to the virtual ground voltage V_- which results in

$$V_{S,\phi 2} = -V_{OS} + \frac{1}{A} V_{int} \tag{4.89}$$

$$\approx -V_{OS} \qquad (A \gg 1) \tag{4.90}$$

[21] Charge injection effects due to switching were neglected.

if the offset voltage is assumed constant during both phases. A charge associated with an offset-free voltage ΔV_S given by

$$\Delta V_S = V_{in} + \frac{1}{1+A} V_{OS} - \frac{1}{A} V_{int} \quad (4.91)$$

$$\approx V_{in} \quad (4.92)$$

is then transferred to the capacitor C_I, leading to the desired relation between V_{in} and V_{int}. A problem of such scheme is that during phase $\phi 1$ the amplifier output is pulled to V_{OS} and the OPAMP must have a high slew rate and fast settling time to enable V_{int} to slew back and forth at each clock transition. But its simplicity (and expected low-power consumption characteristic) and offset-cancellation effectiveness justified its choice.

A more serious problem, however, occurs if the amplifier has a finite DC open-loop gain A, which results in a non-zero differential voltage v_d at the input of the OPAMP, when in practice offset cancellation is not ideal and a residual offset appears[22]. According to (4.91), in the presence of a typical offset of ±5mV [78], a DC gain larger than 65dB is needed so a residual offset smaller than 3μV is achieved[23]. This shows the importance of an amplifier with a very high gain so to guarantee that errors due to offset (and other errors introduced after the first integrator when referred to the input) are negligible. Charge injection though, as mentioned, after offset cancellation is one major limiting factor of performance.

4.8 System level chopping

Dynamic offset-cancellation techniques can be distinguished in two main groups: autozeroing and chopping. The fundamental difference between them is the way the offset is handled. While in the autozeroing technique the offset is sampled first and subtracted in the next phase (what makes this technique appropriate to sampled systems), the chopping technique modulates the offset to higher frequencies, so it can be removed by low-pass filtering [24]. Despite of differences, these techniques and their variants, e.g. CDS, ping-pong, chopper-stabilization, etc., have the same functionality greatly succeeding in canceling the offset.

However, whether autozeroing or chopping, offset residuals often appears. They may have different causes, depending on the technique used, but the

[22] Leakage, given by the last term in (4.91) has also a negative effect [22].

[23] A more conservative condition for the OPAMP DC gain of a B-bit resolution A/D converter is given by $A_{dB} = 20log(2^B)$. For $B = 15$ bits, a DC gain larger than $90dB$ is needed.

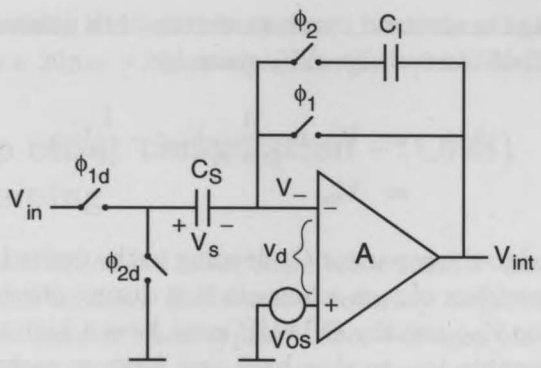


Figure 4.20: Autozeroing scheme used to cancel the offset of the first integrator of the modulator [43]. Clock phases ϕ_1 and ϕ_2 are non-overlapping.

consequence is the same: after offset cancellation, there is still a remaining offset that, depending on the application, can degrade the system performance. In the particular case of this work, an offset voltage of tens of μV is still not acceptable, since inaccuracies due to offset below 0.01°C are only achieved with offset voltages bounded to some units of μV. Some further offset cancellation is then needed.

Some techniques like Nested-chopping [24][79], Three Signal [7] and other system level chopping techniques have been reported as further offset- (as well as charge injection) cancellation methods, since they in general handle the common cause of these residuals: switches mismatch.

The same system level chopping as reported and implemented in [22] was used in this work too. The main reasons behind such choice are the error cancellation levels it provides, effectiveness that is fundamental for high accuracy applications, and mostly important the implementation simplicity, suited for $\Delta\Sigma$ modulators without increasing power consumption.

Its operation can be explained referring first to a chopped autozeroed integrator, as shown in 4.21, which is the same of a chopped amplifier, except for the state inversion (capacitors switching) the integrator needs [70].

The residual offset $V_{OS,res}$ that eventually appears in one integration capacitor, e.g. C_{I1b}, due to the offset V_{OS} and other mismatching errors (charge injection and leakage), after a half conversion when $chop = 0$, is canceled by a residual offset that will eventually accumulate in the other capacitor C_{I1a}(whose position is exchanged with C_{I1b}) during the remaining conversion when $chop = 1$. Since chopping of an autozeroed integrator occurs at a much lower frequency than the autozeroing operation, errors due to charge injection are considerably smaller (e.g. typically a 100 times smaller)

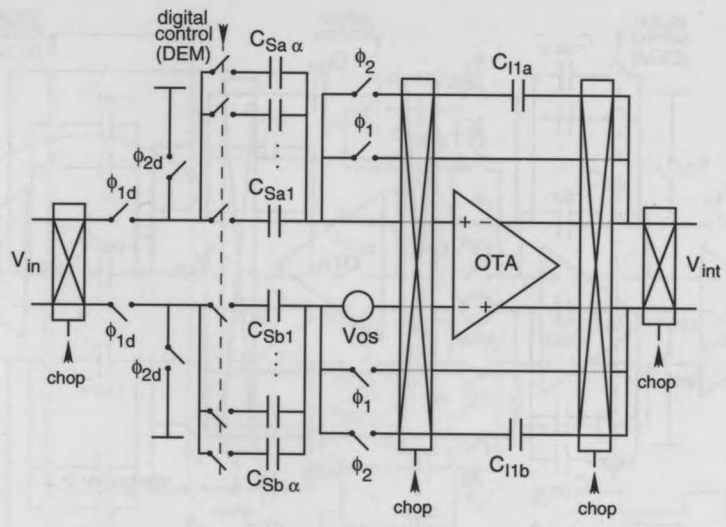


Figure 4.21: Schematic of a chopped autozeroed switched-capacitor integrator.

and the residuals are then expected to be below μV.

System level chopping operates in a very similar way. However, instead of having the last chopper right after the first integrator, this chopper is placed at the output of the modulator, after the comparator, as shown in Fig. 4.22. The second (and other integrators, if present) must have choppers too, so to perform the state inversion as needed. In this case, other errors that might occur inside the modulator, not necessarily in the first integrator, are also canceled.

4.9 ΔΣ modulator implementation

Figure 4.23 shows the schematic of ΔΣ modulator (CIFB-IF) of the sensor's A/D converter. Two differential OTAs implement the switched-capacitor integrators (in cascade) that together with a track-and-latch comparator compose the modulator. The input is feed-forwarded to the second integrator (refer also to [8][22]).

Because of the importance of the first integrator in the modulator's overall performance, special care was given to it. This includes dynamic offset cancellation techniques (autozeroing and chopping at system level), DEM and a very precise analog design at the circuit level, despite of other techniques used to increase performance at the circuit level, e.g. gain boosting [80]. In this work, efforts are done in the sense to keep the benefits the ar-

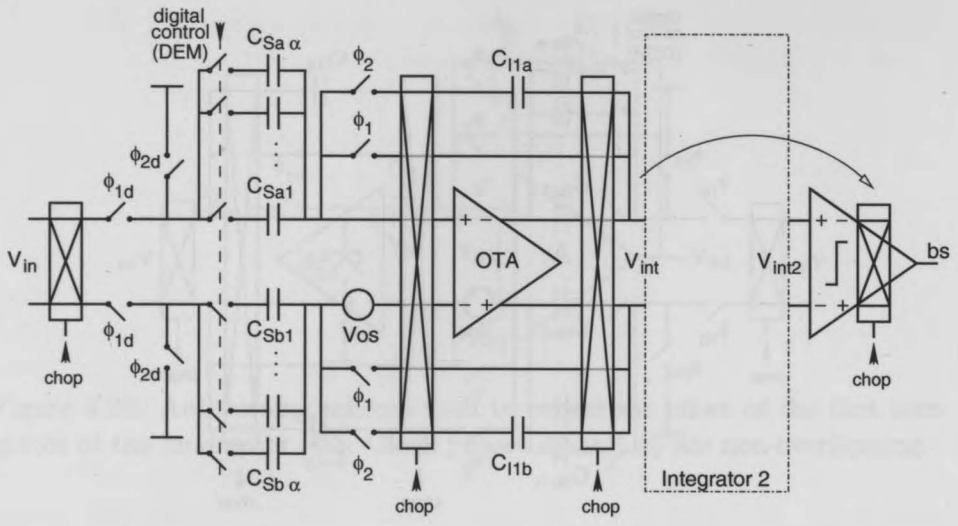


Figure 4.22: Schematic of the second-order ΔΣ modulator with system level chopping in detail. The modulator input V_{in} (either ΔV_{BE} or V_{BE}) is the front-end output (chopped second integrator is not shown).

chitecture provides in terms of accuracy, while reducing the power consumed to meet the the sensor's specifications.

4.9.1 Integrators

Fundamental requirements

Bandwidth and DC gain are the most important parameters of an Operational Transconductance Amplifiers (OTA) to be used in a SC integrator [67][22]. With regard to bandwidth, the amplifier loaded with all the capacitive paths has to settle to the required accuracy in the available time. In terms of DC gain, a value high enough, under all operating conditions and process corners, is necessary to guarantee negligible leakage in the first integrator and to ensure that errors introduced after it can also be neglected.

Another important requirements concern power consumption and noise. This is particularly important in this type of application, a highly accurate temperature sensor. Given the same g_m, a simple OTA, i.e. with a small number of current branches, often translates to a lower-power design. This also often means a smaller number of transistors contributing to the amplifier noise, so the total input referred white noise of the OTA is reduced.

Other requirements usually apply. They are normally related to input

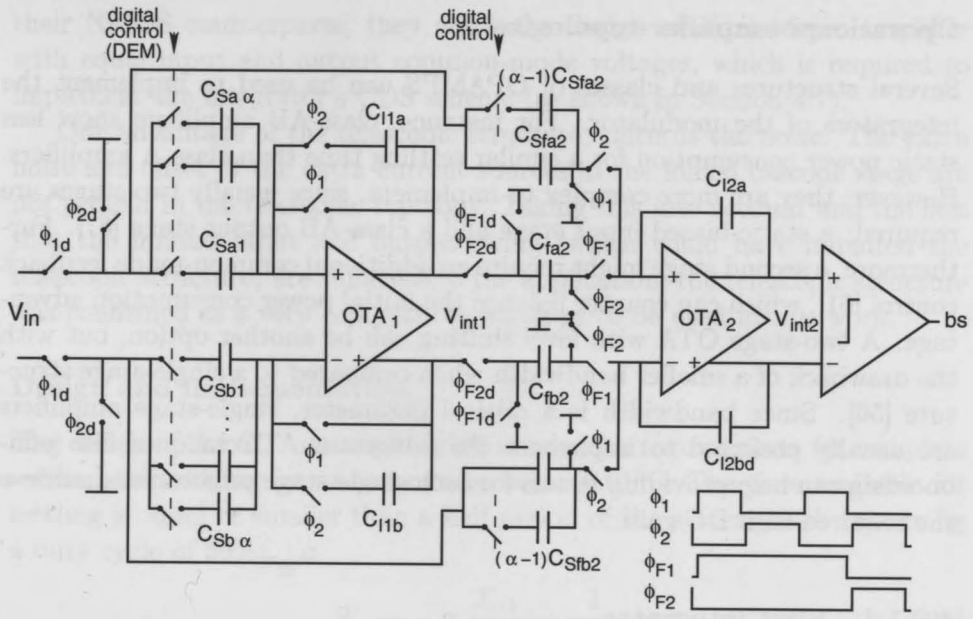


Figure 4.23: Schematic of the second-order ΔΣ modulator (CIFB-IF). The modulator input (either ΔV_{BE} or V_{BE}) is the front-end output (system level chopping is not shown).

Table 4.4: OTA main target specifications for the ΔΣ modulator's first integrator.

Parameter	Value
DC gain (dB) over corners	≥ 100
GBW= $\frac{g_m}{2\pi C_L}$ @ $C_L = 10$pF (kHz)	≥ 200
phase margin (°)	≥ 60
input&output common-mode range (V)	1.2
supply voltage (V)	$2.5 - 5.5$
total supply current (μA)	≤ 5.0

and output common-mode voltages, input and output range, among others. Depending on the application, they have to be checked for compliance with the amplifiers specifications.

Table 4.4 summarizes the main target specifications of the OTA used to implement the first integrator of the ΔΣ modulator.

Operational amplifier topologies

Several structures and classes of OPAMPS can be used to implement the integrators of the modulator. For instance, class-AB amplifiers show less static power consumption for a similar settling time than class-A amplifiers. However, they are more complex to implement, since usually two stages are required: a static-biased input stage and a class-AB output stage [67]. Furthermore, a second stage might require an additional common-mode feedback control [61], which can counter balance the initial power consumption advantage. A two-stage OTA with level shifting can be another option, but with the drawback of a smaller bandwidth when compared to a single-stage structure [56]. Since bandwidth is a critical parameter, single-stage amplifiers are usually preferred to implement the integrators. Techniques like gain-boosting can help providing means for such single-stage structures to achieve the required high DC gain.

4.9.1.1 First integrator

The first integrator is built around a gain-boosted telescopic OTA with a PMOS input pair, as shown in Fig. 4.24. Compared to the more commonly used folded-cascode topology, the telescopic topology offers better power efficiency at the expense of reduced input and output swing and higher minimum voltage supply [47]. Gain-boosting [81][80] is required to achieve a DC gain high enough, over all process corners and operating conditions, to ensure negligible errors due to sampling and leakage.

Since the telescopic structure only needs half the supply current of the folded version for the same maximum output current and g_m, a noticeable power saving is possible. This makes such structure very attractive for low power designs. A disadvantage, though, is a minimum supply voltage one saturation voltage higher than the folded version. Another drawback, since input and output now share the same branch, is a very limited output swing, not to mention that the common-mode input voltage cannot cover much of the output voltage range. For a full-scale input-signal excursion, this would require smaller modulator's scaling coefficients, so to avoid overloading the integrators, which in practice means a smaller conversion resolution for the same clock periods [63]. However, the output limited swing is not an issue in this particular application, due to the characteristics of the input signal that uses roughly only 30% of the modulator's input range. With regard to the input and output common-mode levels, this was effectively addressed by using PMOS transistors (with high threshold transistors, available in the targeted technology) in the amplifier's input pair. Although not as fast as

their NMOS counterparts, they allow the design of a telescopic amplifier with equal input and output common-mode voltages, which is required to implement the integrator's CDS scheme (as shown in Section 4.7).

One advantage of the telescopic structure concerns the noise. The extra noise and offset of the extra current sources in the folded cascode stage are not present in the telescopic topology. Taking this into account and the fact that the limited input and output swing (which could have impaired the telescopic structure) are sufficient to the application, the telescopic structure was confirmed as a very competitive topology to be used in this work.

Design and implementation

The design of the OTA starts from its bandwidth requirements for complete settling, i.e, settling for an inaccuracy below 0.01°C. The time available for settling is equal or smaller than a half-period of the system clock (assuming a duty cycle of 50%), i.e.

$$T_{settling} \approx \frac{T_{clk}}{2} = \frac{1}{2f_{clk}}. \tag{4.93}$$

where f_{clk} is the operating frequency.

The settling time constant τ, during the integration phase, which determines the settling speed of the integrator, is given by

$$\tau = RC_{cl} = \frac{C_{cl}}{g_m} \tag{4.94}$$

where g_m is the integrator's transconductance and C_{cl} the closed-loop capacitance determined from the open-loop capacitance divided by the feedback factor β_{fb}, i.e. $C_{cl} = C_{ol}/\beta_{fb}$ (refer to Appendix F). If a settling of B bits is required, then a number $N_\tau = \ln(2^B)$ of time constants τ is necessary, i.e.

$$T_{settling} = N_\tau \tau = N_\tau \frac{C_{cl}}{g_m}. \tag{4.95}$$

The input transistors of the OTA were designed with a large g_m/I_D ratio to push the gain-bandwidth of the amplifier, loaded with C_{ol}, to the value imposed by settling requirements with a minimal drain current I_D, i.e. the input transistors were designed with a large aspect ratio to push them in weak inversion[24]. In this region, the transconductance g_m over the drain current I_D is approximated to (refer to Appendix E)

$$\frac{g_m}{I_D} = \frac{1}{nV_T}, \tag{4.96}$$

[24]Since the input capacitance C_P should be minimized (or the gate area given by $W \cdot L$), a small channel length is normally used.

where $n \approx 1.5$ is the slope factor. Combining this result with (4.93) and (4.95) gives

$$I_D = \frac{\ln(2^B) 2 f_{clk} C_{cl}}{\frac{g_m}{I_D}} \tag{4.97}$$

which relates the minimum biasing given the amplifier's load capacitances.

For a $\frac{g_m}{I_D} \approx 20$ (weak inversion), $f = 10\text{kHz}$, $B = 15\text{b}$, $C_{cl} = C_{ol}/\beta \approx 28\text{pF}$ ($C_S = \alpha C_{S,min} = 20\text{pF}$, $C_I = 10\text{pF}$ and $C_L = 2\text{pF}$), a minimum $I_D = 0.3\mu\text{A}$ is required flowing through the input pair transistors of the OTA[25]. In order to have a certain margin for a complete settling (but also because simulations over all corners and operating conditions have revealed this need), a transistor current $I_D = 0.75\mu\text{A}$ was used, which gives the OTA a tail current of $1.5\mu\text{A}$. A current of $I_D = 0.75\mu\text{A}$, which is 3 times the front-end's bias current, assures that settling time is then mainly determined by the front-end's current. Although a similar bias in both circuits would be more power efficient due to a more balanced driving capability, a guaranteed performance of the amplifier was only achieved with the increased bias. Figure 4.24 shows the gain-boosted fully-differential telescopic-cascode OTA.

Gain-booster amplifiers

In general terms, a single-stage cascode amplifier provides no more than 80dB of DC gain, which is not sufficient for this particular application. In order to achieve a DC gain well above 100dB with a only one stage, gain-boosting is required. The gain-booster OTAs are shown in Fig. 4.25. They are fully differential folded-cascode OTAs. The input common-mode voltages for both were derived from the tail node from the main amplifier. The main care was to limit the power consumption, which is in general easily achieved since these amplifiers do not need to be fast compared to the main amplifier. Attention was paid to the settling characteristic of the main OTA which required care with the doublet [80] (small compensation capacitors $C_C = 500\text{fF}$ were used).

The switched-capacitor common-mode feedback control

The common-mode voltage at the input of the amplifier is set by its output (both input and output common-mode voltages have the same value, 1.2V) during the sampling phase, when the amplifier is in unity-gain feedback. The output common-mode voltage, however, has to be controlled inside the OTA. This is accomplished by the common-mode feedback-control circuit, whose

[25] This result agrees with that presented in (4.74) if V_T is used instead of $\frac{g_m}{I_D}$ and $B = 14\text{b}$, which is equivalent to an error $\varepsilon = 0.0067\%$.

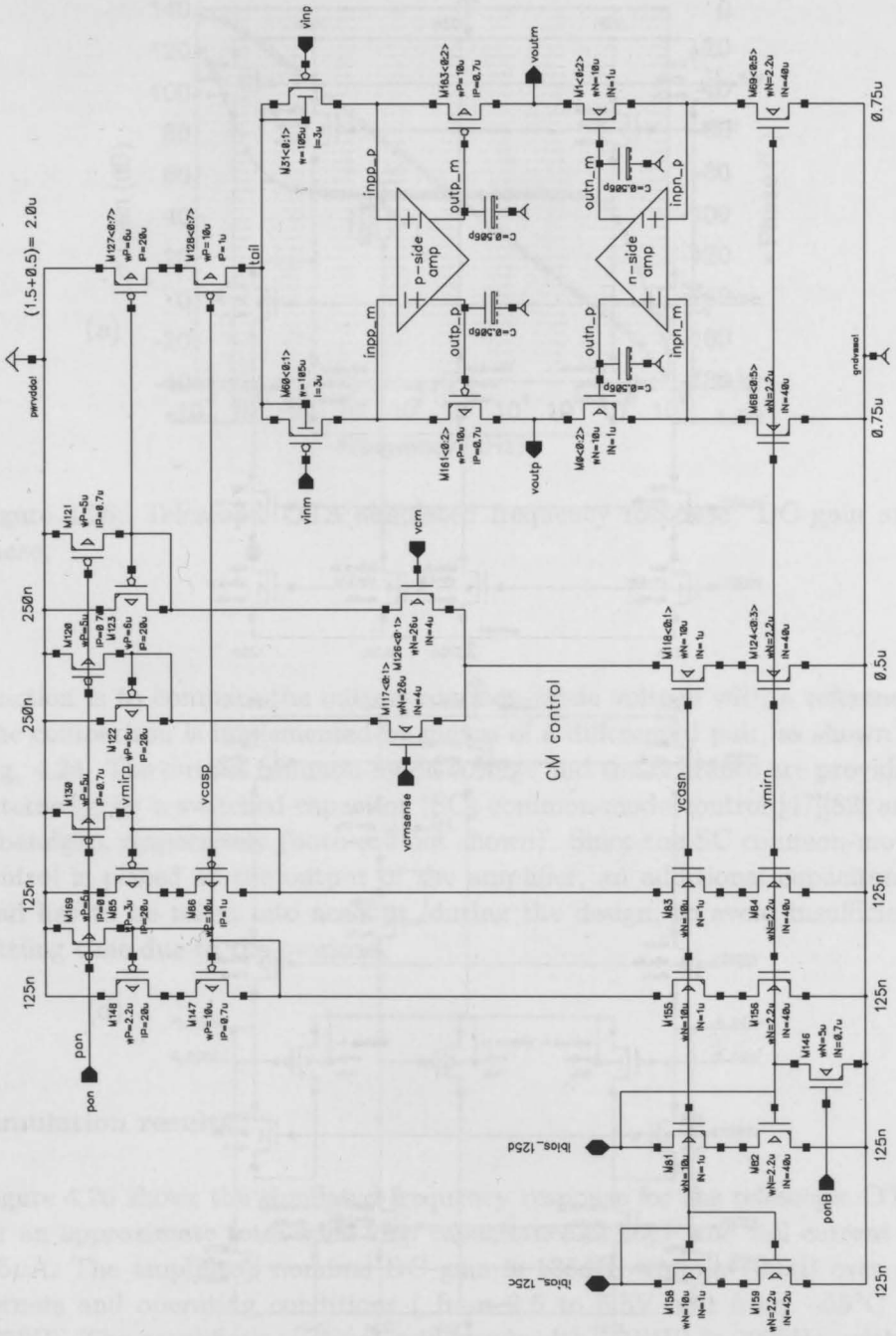


Figure 4.24: Schematic of the fully-differential gain-boosted telescopic-cascode OTA used to implement the first integrator.

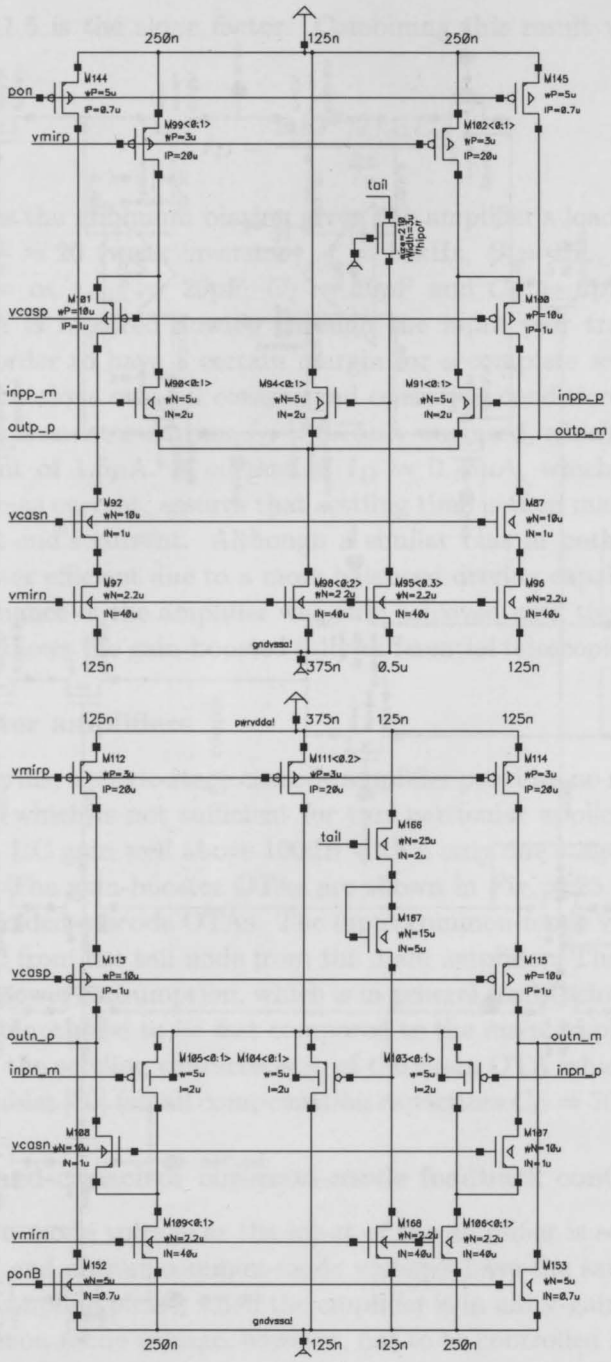


Figure 4.25: Schematics of the gain-booster amplifiers. (a) p-side amplifier; (b) n-side amplifier.

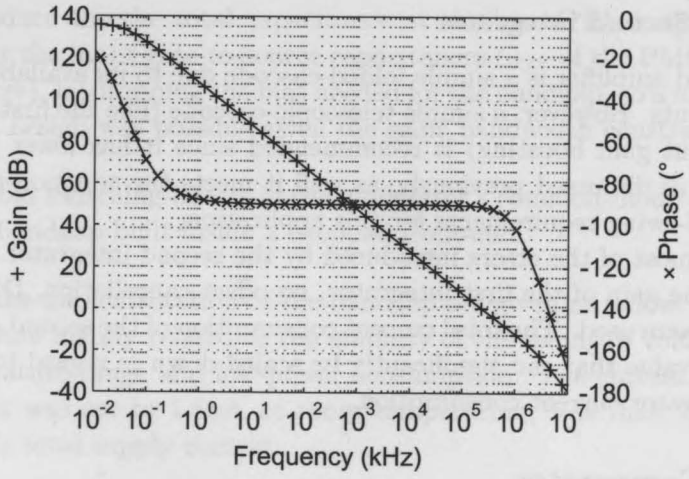


Figure 4.26: Telescopic OTA simulated frequency response: DC gain and phase.

function is to compare the output common-mode voltage with a reference. The comparison is implemented by means of a differential pair, as shown in Fig. 4.24. The output common-mode voltage and the reference are provided externally, by a switched-capacitor (SC) common-mode control [47][82] and a bandgap, respectively (both are not shown). Since the SC common-mode control is placed at the output of the amplifier, an additional capacitance load has to be taken into account, during the design, to avoid insufficient settling time due to the overload.

Simulation results

Figure 4.26 shows the simulated frequency response for the telescopic OTA for an approximate total open-loop capacitance of 10pF and tail current of 1.5μA. The amplifier's nominal DC gain is 135dB, while a 100dB over all corners and operating conditions (from 2.5 to 5.5V and from -55°C to 125°C). The amplifier's nominal gain-bandwidth (GBW) is 200kHz, with a phase margin of 60°. The input and output common-mode level is 1.2V, for a 800mV voltage swing at the output. Including the boosting amplifiers and biasing, the OTA draws a total current of 5μA.

4.9.1.2 Second integrator

The second amplifier is a simple folded-cascode due to its availability during the tape-outs. However, a simple telescopic-cascode (like the first integrator, but without gain boosting) is recommended since it has lower power consumption as discussed previously, as well it meets the common-mode levels and signal-swing requirements for this application.

Since most of the errors introduced by the second integrator are attenuated by the gain of the first integrator, no offset cancellation, DEM or gain boosting were used. The total current consumption of the second comparator is 3μA, a value that can significantly be scaled down in regard to the actual first integrator current consumption.

4.9.2 Comparator

The used comparator is implemented as a track-and-latch stage, preceded by a pre-amplifier [22]. The underlying reason behind this architecture is as follows: the pre-amplifier is used to obtain higher resolution and to minimize the effects of kickback[26] to the output of the second integrator, which may result in limited accuracy [60]. The output of the amplifier is still much smaller than the voltage levels needed to drive the digital circuitry. The track-and-latch stage then amplifies this signal further during the track phase, and again during the latch phase, when the positive feedback is enabled. The positive feed-back generates the analog signal into a full-scale digital signal. The total current consumption of the comparator is 5μA. The response time is 1μs. While no scaling is applicable to the comparator (as expected to the integrators), a specific design for the comparator might help reducing the modulator's power consumption without any loss of performance.

4.9.3 Bandgap

The $\Delta\Sigma$ modulator does not require any explicit reference voltage, since this reference is built internally by the charge balancing mechanism, as explained. However, a stable reference voltage or bandgap is required to provide the common-mode level to the integrators of the modulator. Even though this circuit has a similar structure to that of the front-end, the bandgap was designed apart from it. The main reasons were:

- to remove from the front-end any other function not related to V_{BE} and ΔV_{BE} generation, keeping in it only the sensor's most critical functions;

[26] Charge transfer either into or out of the driving circuitry when the track-and-latch stage goes from track to latch mode.

- to reduce the the total capacitance at the bias OTA output (largely set by the input gate-to-source capacitances C_{GS} of the PMOS current sources), by dropping the bias needed for the bandgap, so a lower-power OTA design was possible given the same bandwidth constraints;

- to avoid switching-effects in the modulator's common-mode signal, due to a bandgap built inside a chopped front-end.

Details of the implementation of the bandgap are given as follow. The specifications were mainly related to the stability of the bandgap voltage within the temperature range and low-power consumption. The current budget to such circuit was set to 1.5μA at room temperature, less than 10% of the modulator's total supply current.

Design and implementation

The bandgap schematic is shown in Fig. 4.27. It is composed of a core, defined by transistors M_0, M_5, M_7, M_9 and resistor R_1 that provides a PTAT current I. This current is used to bias a PNP transistor in series with a resistor R_2, so a temperature-independent voltage V_{bg} is established.

Two separate identical current sources, defined by the transistors M_0 and M_5 in saturation, provide the current I to transistors M_7 and M_9. These current sources are controlled simultaneously by the amplifier (M_6) in such a way that the current sources operate under equal bias conditions. The influence of the finite output impedance of the devices in the bias current I is thus essentially eliminated [83].

Transistors M_7 and M_9, and the resistor R_0 define the following equation:

$$V_{R0} + V_{GS7} - V_{GS9} = 0 \tag{4.98}$$

The current in the transistor M_7 (or M_9), both operating in weak inversion is given by (refer to Appendix E)

$$I_{D7(9)} = I_0 \left(\frac{W}{L}\right)_{7(9)} \exp\left(\frac{V_{G7(9)}}{nV_T}\right) \exp\left(\frac{-V_{S7(9)}}{V_T}\right). \tag{4.99}$$

Since $V_{S7} = V_{R0}$, $V_{S9} = 0$, $V_{G7} = V_{G9}$ and $I_{D7} = I_{D9}$ (enforced by the current mirrors), it is straight forward to show that

$$V_{R0} = V_T \ln\left(\frac{\left(\frac{W}{L}\right)_7}{\left(\frac{W}{L}\right)_9}\right) \tag{4.100}$$

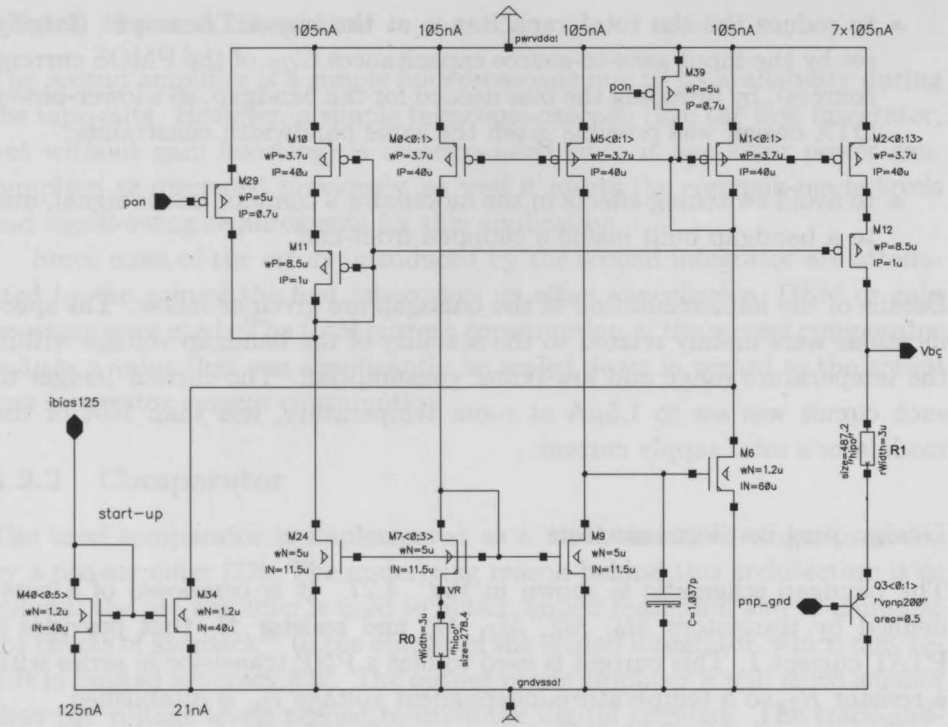


Figure 4.27: Bandgap schematic.

or $V_{R0} = V_T \ln(4)$ if $\left(\frac{W}{L}\right)_7 = 4 \cdot \left(\frac{W}{L}\right)_9$. This voltage, when applied to the resistor R_0 sets the PTAT current $I_{PTAT} = V_T \ln(4)/R_0$. A current $I_2 = k \cdot I_{PTAT}$ is then used to bias the PNP transistor, so a voltage V_{BG} is established:

$$V_{BG} = -V_{BE} + I_2 R_1. \tag{4.101}$$

Because (4.101) is a sum of a CTAT voltage with a PTAT voltage, it is at first order temperature independent[27].

Simulation results

Figure 4.28 shows the bandgap simulation results. The voltage $V_{BG} \approx 1.27$V (as shown in Fig. 4.28(a)) is within ± 1.5mV from $-55°$C to $125°$C. Since this voltage is not chopped, simulation results have shown a modulator with a settling characteristic much more similar to a first order system. Moreover,

[27] The exact ratio R_1/R_0 to achieve such characteristic can be determined by simulation or by taking $\partial V_{BG}/\partial t = 0$ at the temperature of interest. The advantage of using a simulator is that it takes into account second- and higher-order effects.

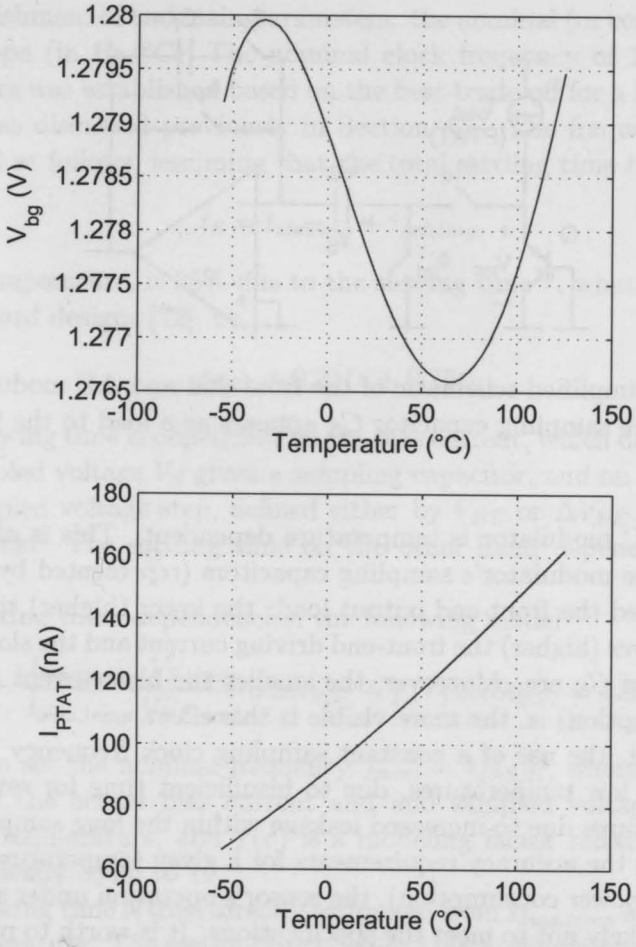


Figure 4.28: Banddap simulated DC response (temperature sweep). (a) Bandgap voltage V_{BG} and (b) PTAT bias current I.

the common-mode voltage built from a bandgap gives to the A/D converter insensitivity to oscillations in the supply voltage.

4.10 PTAT clock frequency

The front-end bias current is PTAT. As discussed in Chapter 3, biasing the PNP transistors with a PTAT current reduces the curvature of V_{BE}. This allows to build the reference voltage V_{REF} with a smaller non-linear component, so more accurate measurements can be achieved.

However, because of the front-end PTAT bias current, the total settling-

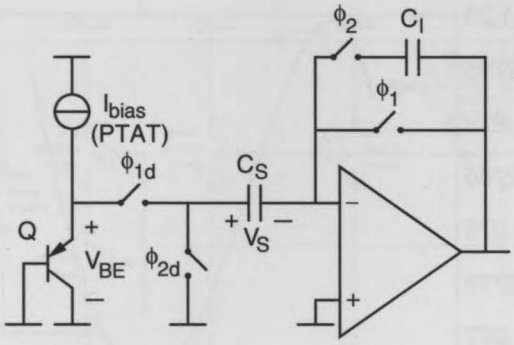


Figure 4.29: Simplified schematic of the front-end and ΔΣ modulator's first integrator. The sampling capacitor C_S appears as a load to the front-end.

time of the SC modulator is temperature dependent. This is easily understood since the modulator's sampling capacitors (represented by C_S in Fig. 4.29) are indeed the front-end output load: the lower (higher) the temperature is, the lower (higher) the front-end driving current and the slower (faster) the charging of C_S are. Moreover, the smaller the bias current (for a lower power consumption) is, the more visible is this effect.

As a result, the use of a constant sampling clock frequency leads to errors either at low temperatures, due to insufficient time for settling, or at high temperatures due to increased leakage within the long sampling period. Depending on the accuracy requirements for a given temperature range and bias current (power consumption), the sensor's operation under a fixed clock frequency is likely not to meet the specifications. It is worth to mention that the PTAT current itself does not cause this effect, but together with a low bias current, which introduces a slewing time not negligible when compared to the total settling-time.

In order to minimize such errors, the sensor uses a sampling clock with a temperature-dependent frequency. At low temperatures, the frequency is lower than its nominal value, allowing more time for settling, while at high temperatures, it is greater, reducing leakage errors by shortening the sampling period. The sensor, despite of having a conversion time that is now temperature dependent[28], maintains its accuracy over the full temperature range.

The success of this approach, however, is strongly dependent on the cor-

[28]This is easily solved at the system level (by realizing that the same happens with the sensor at the circuit level when operating with a fixed clock frequency) by defining a measurement time based on the worst case conversion time, i.e. at the lowest temperature.

rect establishment of two main parameters: the nominal (or center) frequency and its slope (in Hz/°C). The nominal clock frequency of 11kHz at room temperature was established based on the best trade-off for a fixed frequency operation as discussed previously in Section 4.5. The frequency slope was established as follows, assuming that the total settling time t_S given by

$$t_S = t_{slewing} + t_{settling}, \tag{4.102}$$

at room temperature, is 25% due to the slewing time[29], what usually agrees with standard designs [72], i.e.,

$$t_{S,27} = 0.25t_S + 0.75t_S \tag{4.103}$$

The slewing time is dependent on the bias current, which defines the slope of the sampled voltage V_S given a sampling capacitor, and on the magnitude of the sampled voltage step, defined either by V_{BE} or ΔV_{BE}, depending on the $\Delta\Sigma$ cycle. The settling time on the other hand depends on the time constant τ.

Accounting these dependencies, the following model

$$t_{S,temp} = \frac{I_{bias,nom}}{I_{bias,temp}} \frac{V_{S,temp}}{V_{BE,nom}} 0.25t_{S,27} + f(\tau)\, 0.75t_{S,27} = 1.0t_{S,27} \tag{4.104}$$

was used to set the nominal frequency $f_{nom} = 1/t_{S,27}$, where $I_{bias,temp}$ and $V_{S,temp}$ are the actual bias current and and sampled voltage (e.g. V_{BE}) at a given temperature, and $f(\tau)$ is a modeling factor relative to the time constant, ideally equal to 1.

The slewing time is then directly estimated from $I_{bias,temp}$ and $V_{S,temp}$ at a given temperature. The settling time, which ideally should be approximately constant, presented a variation of ±20%, according to simulations, mainly due to the variation of the PNP transistor's output impedance. The switches resistance R_{on} were not modeled.

For a front-end nominal PTAT bias current $I_{nom} = 250$nA with a minimum value of $0.63I_{nom}$ at −55°C and a maximum value of $1.54I_{nom}$ at 125°C, and a nominal voltage $V_{BE,nom}$ of 0.55V at room temperature, with a maximum value of 0.73V at −55°C and minimum value of 0.32V at 125°C, the estimated total settling time, at low temperature, is

$$t_{S,-55} = \frac{1}{0.63}\frac{0.73}{0.55} \cdot 0.25t_{S,27} + 1.2 \cdot 0.75t_{S,27} = 1.42t_{S,27} \rightarrow 0.7f_{nom}, \tag{4.105}$$

[29]The exact ratio is usually of difficult establishment, but it agrees roughly with the present case for the particular bias current and accuracy levels (that set the number of time constants needed for settling). Specific derivation can be found in [22].

Table 4.5: PTAT clock frequency - System operating clock as a function of temperature (or front-end's bias current).

temp. (°C)	-50	-30	-10	10	(27)30	50	70	90	110	130
freq. (kHz)	7.5	8	9	10	11	12	13	14	15.25	17.5

with respect to the nominal value, while at high temperature is

$$t_{S,125} = \frac{1}{1.54}\frac{0.32}{0.55} \cdot 0.25t_{S,27} + 0.8 \cdot 0.75t_{S,27} = 0.69t_{S,27} \rightarrow 1.4f_{nom}. \quad (4.106)$$

These results agree reasonably with the front-end's bias current variation over temperature. The frequency variation of 50Hz/°C or 1kHz/20°C (except at the high end of the temperature range) as shown in Table 4.5 was called PTAT clock frequency. It does not mean, however, that the frequency should vary accordingly to the bias current, but that, for this particular case, the same variation of the bias current should be applied to the system clock so an equivalent amount of time is available for the converter while processing its input, regardless of the temperature. The PTAT clock frequency is certainly not required if the slewing time is negligible with regard to the settling time nor will give an appropriate result if the slewing time is much greater than the 25% of the total settling time assumed here.

Because of the model's simplicity, measurements with other slopes were performed, but they presented no superior performance. For instance, a smaller frequency variation was not enough to compensate the different settling times due to the PTAT bias. On the other hand, a more strong frequency variation ended causing problems mainly at high temperatures due to the lack of integrator's bandwidth.

Measurement results using the PTAT clock frequency showed the same system performance in the middle of the temperature range, as expected, but presented great improvement at the low and high ends of the temperature range. For instance, compared to the use of a fixed (11kHz) clock frequency, this technique leaded to a 50% reduction in the sensor's batch-calibrated inaccuracy at the low end of the temperature range, proving its efficiency. This is shown in the next Chapter 5.

4.10.1 Implementation of the PTAT clock frequency

The PTAT clock frequency was implemented at the system level, by setting the modulator's clock frequency accordingly to the temperature of the thermal chamber (used for measurements). In practice, the implementation of a PTAT clock frequency oscillator will rely in the design of a Voltage Con-

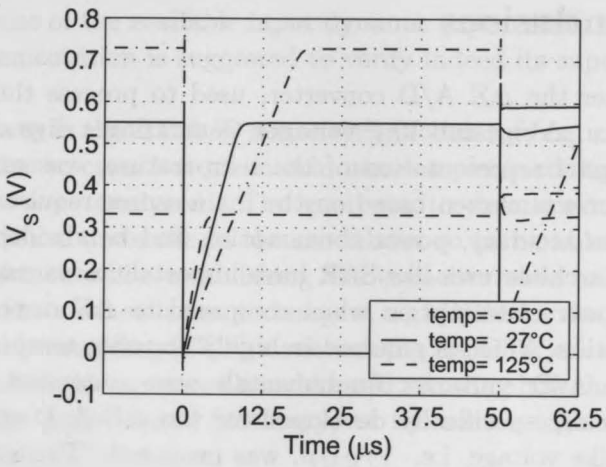


Figure 4.30: Slewing time dependence on temperature (during a V_{BE} cycle, $f = 10\text{kHz}$). The time remaining for settling, that sets the sensor's accuracy, is a strong function of temperature.

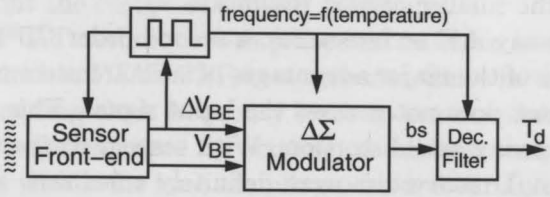


Figure 4.31: A simple implementation of the clock generator (whose frequency is PTAT). The VCO control signal is the PTAT voltage ΔV_{BE} generated in the front-end.

trolled Oscillator (VCO) whose control voltage might be efficiently derived from the PTAT voltage ΔV_{BE}, like depicted in Fig. 4.31.

Due to the low-frequency system operation, the implementation of a PTAT clock frequency oscillator has relatively low complexity and might require negligible extra power consumption compared to that of a simple oscillator. In addition, VCOs can be implemented with relaxation and ring oscillators. Since the sensor is based on a SC modulator, a ring oscillator could be used, since phase noise is not important for accuracy but settling. The relaxation oscillator, however, may allow better tuning and controlability through ΔV_{BE} signal. Regardless of the case, this topic is out of the scope of this work.

4.11 Conclusions

In this chapter the ΔΣ A/D converter, used to process the analog front-end output, i.e. ΔV_{BE} and V_{BE} voltages, generating a digital signal, whose value is a digital representation of the temperature, was exposed. A ΔΣ A/D converter was chosen based on the main sensor requirements, initially described, like accuracy, power consumption and bandwidth. While some Nyquist-rate architectures like SAR have important advantages in terms of speed and power consumption when compared to ΔΣ converters, there is lack of resolution, which is required in highly accurate temperature sensors.

After the A/D converter fundamentals were presented, an optimized charge balancing, specifically developed for the ΔΣ A/D converter, based on a PTAT-like voltage, i.e. V_{TEMP}, was proposed. The main motivation was to use more efficiently the available A/D converter input range, relaxing the A/D resolution requirements, with a further and natural reduction of the conversion time. Although the new voltage V_{TEMP} is feasible, its implementation is more complex than the implementation of the classical and well-known PTAT voltage ΔV_{BE}: the longer ΔΣ cycle required now cancels the benefits of the smaller needed resolution.

Despite of many ΔΣ architectures, a second-order CIFF-IF architecture was chosen. One of the major advantages of such architecture is the fact that the first integrator does not process the input signal. This reduces the integrator's output swing and distortion. With respect to the modulator order, first-order ΔΣ A/D converters were definitely ruled out, since the number of cycles needed for a given resolution is too high, e.g. 32768 cycles for 15 bits. For a 10Hz sensor bandwidth, a sampling frequency of about 328kHz is needed, which means a prohibitive integrator's bandwidth in terms of supply current (power consumption). A second-order A/D converter, chosen for this work, requires in theory approximately only 256 cycles for the same resolution (and a clock frequency of about 2.6kHz for the same sensor bandwidth). The increased performance, however, occurs at the expense of a more complex circuitry.

A third-order ΔΣ A/D converter was investigated with no further circuit implementation. Because the first integrator in a second-order modulator is the circuit that more consumes (roughly 65% of the entire modulator's power consumption), the addition of a third integrator should add a small amount to the total modulator's power consumption. On the other hand, the conversion time with a third-order modulator is reduced by about 77% with respect to a second-order modulator for the same resolution due to its more aggressive noise shaping. Remarkably, the smaller available modulator's input range in a third-order modulator fits better the classical V_{PTAT}, and in practice a

more efficient use of the available input dynamic (up to 60%) is observed. A practical implementation is suggested to verify indeed its superior expected performance.

Errors sources in the $\Delta\Sigma$ A/D converter's modulator, e.g. offset and gain errors, errors due to capacitor mismatch, non-complete settling time, charge injection, etc. were analyzed. The goal was to determine accuracy and power consumption trade-offs in the modulator so to be able to set design specifications for it, while establishing the capacitor sizes and integrator's bias currents. Error cancellation techniques, e.g. DEM for device mismatch, correlated double-sampling (CDS) - autozeroing for OTA offset cancellation and system level chopping, were also addressed, showing how these techniques can improve the performance (accuracy) of the sensor with a small or no increase on power consumption.

Based on the results obtained, the circuit design of $\Delta\Sigma$ modulator, comprised of two integrators, a comparator and a bandgap, was presented. Specific details about the design of the amplifier of the first integrator, a fully-differential gain-boosted telescopic-cascode amplifier with a 100dB DC gain, $GBW = 200\text{kHz}$ and a total supply current of $5\mu\text{A}$, were given. Also were given details about the design of a 1.27V bandgap.

Finally the PTAT technique was presented in order to cope with problems caused by the low PTAT PNP bias currents. Since in this situation, the use of a fixed sampling clock frequency leads to errors either at low temperatures, due to insufficient time for settling, or at high temperatures due to increased leakage within the long sampling period, a variable operating clock frequency, adaptively adjusted accordingly to the absolute temperature, was used. Measurement details are given in the next chapter.

CHAPTER 5

Realizations

In this chapter, measurement results of the developed smart temperature sensors are presented. They were all fabricated in a standard 0.7μm CMOS process (AMIS0.7μm [28]) with linear capacitors and high-ohmic poly resistors, at IMEC/Belgium via Europractice, and mounted in ceramic Dual-in-Line (DIL) packages. Because the designs are based on the same architecture, a general system description is briefly given, before going into particular characteristics and measurement results of each design. Design decisions and tradeoff were extensively addressed in the former chapters.

The first design reduced the power consumption of the sensor, as described in [8][22], by reducing the PNP transistors bias current and the $\Delta\Sigma$ modulator's sampling and integration capacitors, so to keep the sensor bandwidth in 10 conversions per second. While the power consumption was lowered, a non-PTAT temperature error characteristic was observed in this sensor. The next design, named micro-power Front-End, addressed this issue with further reducing the power consumption. A temperature sensor with a trimmed inaccuracy (one-point) of ± 0.25°C ($\pm 3\sigma$) from -55°C to 125°C was obtained.

The third (and last) design added to the micro-power front-end a low-power A/D converter (while improving the front-end current mirror performance too). This circuit resulted in a micro-power precision smart temperature sensor with a batch calibrated inaccuracy below ± 0.25°C ($\pm 3\sigma$) from -70°C to 130°C and one-point trimmed inaccuracy below ± 0.1°C ($\pm 3\sigma$) from -55°C to 125°C, with a total supply current of 25μA (at 2.5V and 30°C).

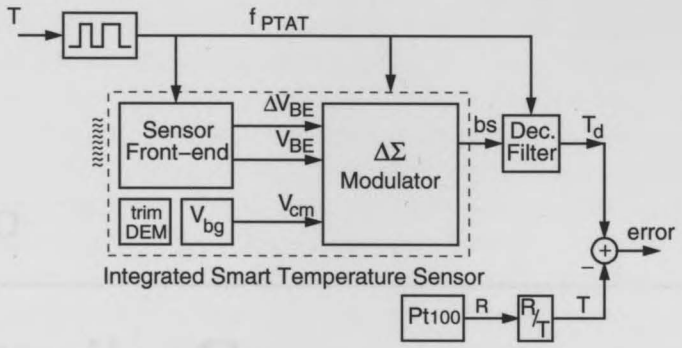


Figure 5.1: Smart temperature sensor block diagram (the Pt100 resistor and R to T converter are not part of the sensor but the temperature error measurement setup used to find the temperature error).

5.1 System description

Figure 5.1 shows a block diagram of the smart temperature sensor. It is built mainly of a (sensor) front-end and a $\Delta\Sigma$ A/D converter. The front-end generates the V_{BE} and ΔV_{BE} voltages which are further processed by the $\Delta\Sigma$ A/D converter. The $\Delta\Sigma$ A/D converter consists of a second-order switched-capacitor modulator and an off-chip decimation filter. A band-gap provides the common-mode voltage to the modulator's differential circuitry. In addition, the PTAT clock generator provides a temperature-dependent operating clock frequency supplied to the front-end $\Delta\Sigma$ A/D converter. Digital circuitry for trimming and DEM control as well other auxiliary circuits are also part of the sensor. A Pt100 sensor used as temperature reference (e.g. needed for calibration) is also shown.

Figure 5.2 shows a more detailed schematic of the smart temperature sensor. The front-end block comprises the bipolar core and the pre-bias circuit. The first generates the voltages V_{BE} and ΔV_{BE} by biasing the PNP transistors with a PTAT current. The PTAT current is generated in the pre-bias circuit with help of an OTA and then mirrored to the bipolar core. The second-order modulator samples these voltages following the charge balancing as explained in Chapter 4, integrating either V_{BE} or ΔV_{BE} depending on the bitstream bs value. In doing this, proper integration gains are set in the first and second sampling capacitors arrays, named C_S and C_{Sf}. DEM and chopping (partially shown) as well autozeroing schemes, as discussed in the last chapter, are also detailed.

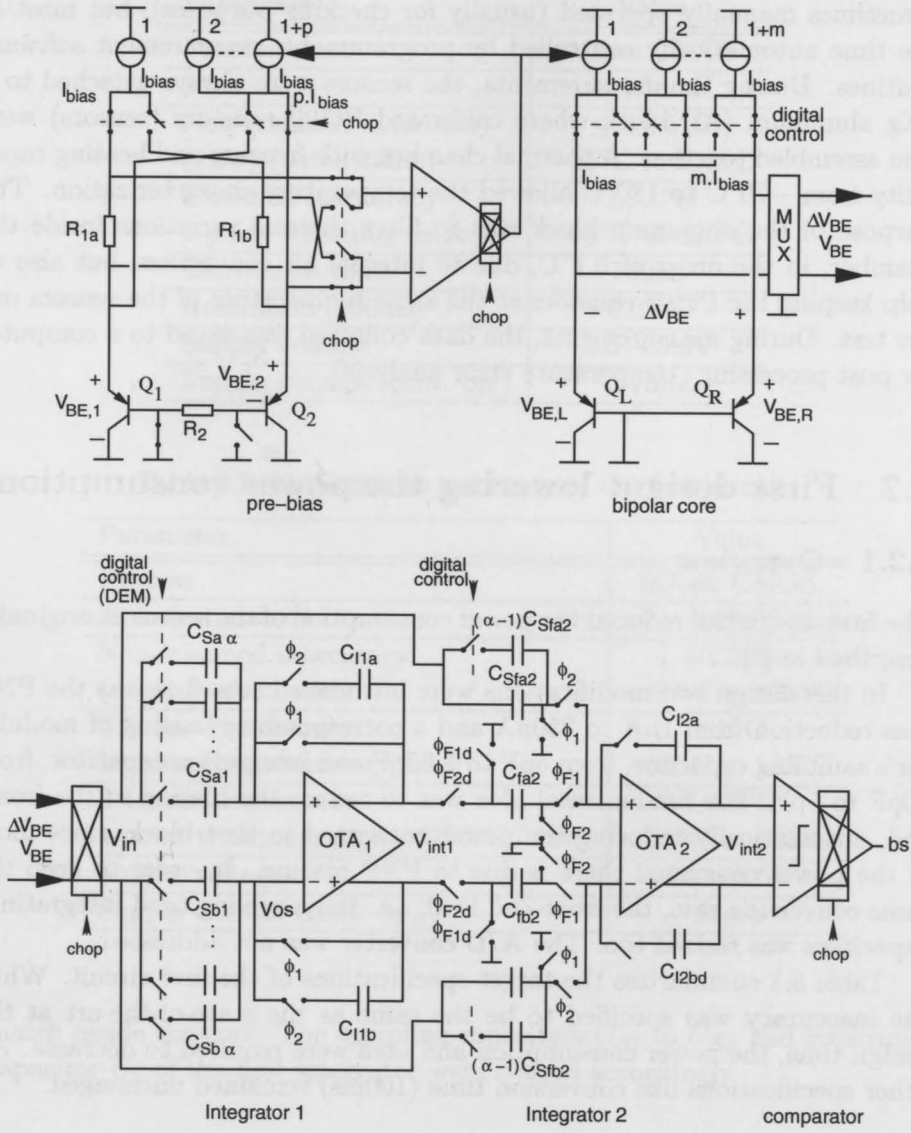


Figure 5.2: Smart temperature sensor circuit diagram. $\phi_1(\phi_{F1})$ and $\phi_2(\phi_{F2})$ are two non overlapping clocks and *chop* is the system chopping control.

5.1.1 Measurement description

All measurements were performed with standard instrumentation equipments, sometimes manually operated (usually for checking purposes), but most of the time automatically controlled by programmable measurement software routines. During the measurements, the sensors were always attached to a 3Kg aluminum (Al) block where calibrated Pt100 resistors (sensors) were also assembled together. A thermal chamber with freezing and heating capability from $-70°C$ to $180°C$ allowed the temperature characterization. The purpose of the aluminum block was to filter thermal variations inside the chamber, in the order of $0.1°C$, due to internal air convection, but also to help keeping the Pt100 resistors at the same temperature of the sensors under test. During measurements, the data collected was saved to a computer for post processing (temperature error analysis).

5.2 First design: lowering the power consumption

5.2.1 Overview

The first design [84] reduced the power consumption of the sensor as originally described in [8].

In this design two modifications were introduced: the first was the PNP bias reduction from $1\mu A$ to 250nA and a corresponding resizing of modulator's sampling capacitor, from 5pF to 1.25pF and integration capacitor, from 20pF to 5pF. The fundamental idea was to reduce the biasing in the front-end, automatically reducing the power consumed in that block, since most of the power consumed there is due to PNP biasing. In order to keep the same conversion rate, the front-end load, i.e. the sampling (and integrating) capacitors was resized too. The A/D converter was not addressed.

Table 5.1 summarizes the target specifications of the first circuit. While the inaccuracy was specified to be the same as the state-of-the-art at the design time, the power consumption and area were required to decrease. All other specifications like conversion time (100ms) remained unchanged.

5.2.2 Implementation

The current in the bipolar core I_{bias} was reduced by decreasing the geometric ratios of its PMOS current sources. No care about matching was taken into account of these resized current sources with those of the pre-bias circuit. The initial belief was that most of the accuracy was relying in the current

Table 5.1: Target specifications summary of the first circuit (based on the performance results of [8][22]).

Parameter	Value
Process	0.7μm CMOS
Area	$< 4.5\text{mm}^2$
Non-trimmed inaccuracy ($\pm 3\sigma$) (batch calib.)	$\leq \pm 0.5$°C -55°C to 125°C
Trimmed inaccuracy 1-point trimming ($\pm 3\sigma$)	± 0.1°C -55°C to 125°C
Conversion time	100ms
Resolution (100ms)	0.01°C
Supply voltage	$2.5 - 5.5$V
Supply current (cont. op.)	$< 75\mu$A

Table 5.2: Performance summary of the first circuit.

Parameter	Value
Process	0.7μm CMOS
Area (estimated)	4mm^2
Non-trimmed inaccuracy ($\pm 3\sigma$) (batch calib.)	± 1.5°C -55°C to 125°C
Trimmed inaccuracy 1-point V_{BE} trimming ($\pm 3\sigma$)	± 1°C -55°C to 125°C
Trimmed Inaccuracy 1-point trimming, digital @30°C ($\pm 3\sigma$)	± 0.3°C -55°C to 125°C
Supply voltage	$2.5 - 5.5$V
Supply current (cont. op.)	55μA

match inside the core. The sampling capacitors C_{S1} to $C_{S\alpha}$ and integration capacitor C_I of the first integrator were resized accordingly.

5.2.3 Experimental results

Table 5.2 shows the performance summary of the circuit. Non-trimmed and trimmed inaccuracy results over temperature are shown graphically as follows.

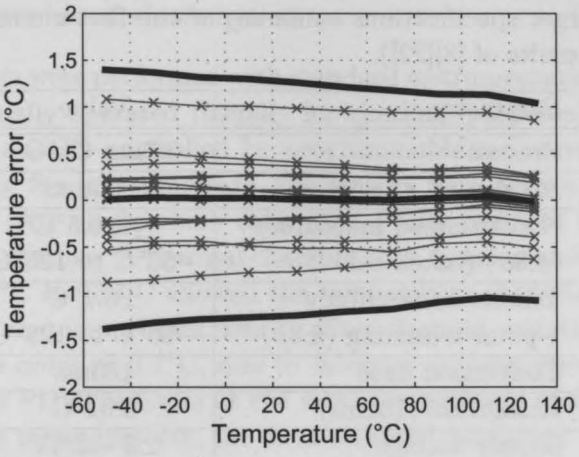


Figure 5.3: Measured temperature error after batch calibration, untrimmed (16 samples with average and $\pm 3\sigma$ limits).

5.2.3.1 Batch calibration results

Figure 5.3 shows the measurement results for 16 devices after batch calibration (untrimmed error), along with the average error and $\pm 3\sigma$ limits. The inaccuracy is ± 1.5°C, mainly due to higher spread between samples when compared to [8]. Also, instead of the expected PTAT spread in inaccuracy due to the PTAT spread of V_{BE}, the data shows a CTAT (complementary proportional to the absolute temperature) error. One explanation is that a voltage $V_{q,inj}$, due to an under-estimated charge injection, added to ΔV_{BE}, modifies μ in (3.8). Because it adds to ΔV_{BE}, and ΔV_{BE} is smaller at low temperatures, it causes a CTAT spread of the error. A second reason is due to the increased current sources mismatch. Because only current sources in the core were resized (i.e. made smaller), poorer match is expected in the core and, much worse, from the pre-bias circuit to the core. This mismatch causes also a CTAT error and worsens the β-compensation scheme. Added together, this may have caused such a large observed temperature errors when compared to [8].

5.2.3.2 Trimmed results

Figure 5.4 shows the temperature error of 8 samples after trimming at 30°C using the V_{BE} method as proposed in [8][27]. Because the untrimmed error has a CTAT behavior, this method is not effective, as discussed in Chapter 2, since it only corrects the PTAT spread of V_{BE}. The resulting error has

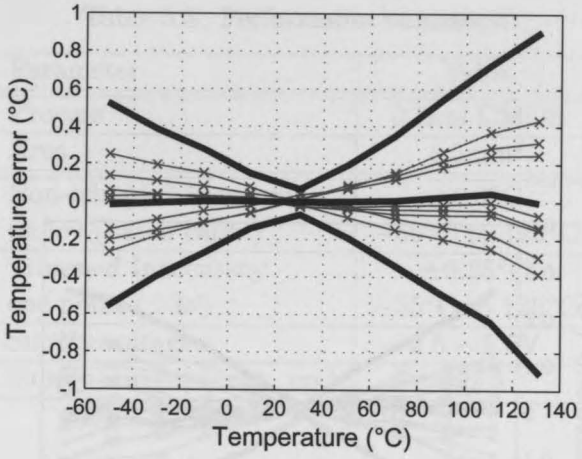


Figure 5.4: Measured temperature error after individual trimming at 30°C (8 samples with average and $\pm 3\sigma$ limits).

then the typical butterfly characteristic.

Better results than those obtained with V_{BE} trimming were obtained with digital trimming which is able to correct PTAT and non PTAT errors. This is shown in Fig. 5.5(a), where an inaccuracy below ± 0.3°C is observed.

It is worth to mention that the former results were obtained after removing the third-order curvature error (that contributes with an error of about ± 0.15°C) and the systematic error due to leakage, at the high-end of the temperature range. If not, the results, for instance, presented in Fig. 5.5(a) become those given in Fig. 5.5(b), where it is clear the third-order curvature and the leakage effect: the inaccuracy increases from ± 0.3°C to approximately ± 0.4°C.

If higher accuracy is needed, a two-point digital calibration is suggested, despite of the higher associated cost. Together with a systematic curvature correction, an error below ± 0.1°C was found over the full temperature range (not shown).

The sensor draws a total current of 55μA (at 3.3V, 30°C). This represents a power reduction of 22% with respect to [8], despite of the poorer accuracy, further addressed in the next designs. Regarding the PNPs bias currents in the bipolar core, the total current was reduced by 75%.

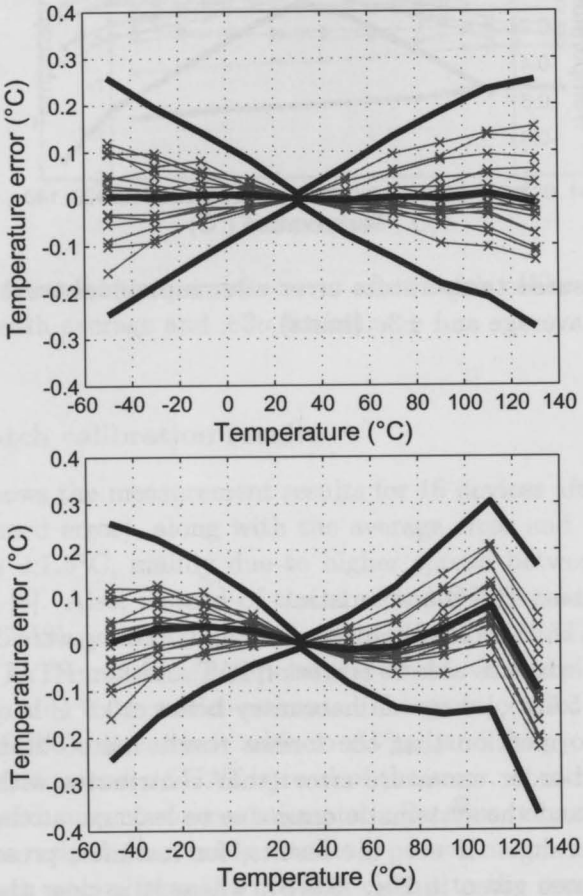


Figure 5.5: Measured temperature error after digital trimming at 30°C (a) with and (b) without third-order curvature correction (16 samples with average and $\pm 3\sigma$ limits).

Table 5.3: Performance summary.

Parameter	Value
Process	0.7μm CMOS
Area	4.5mm^2
Non-trimmed Inaccuracy ($\pm 3\sigma$) (batch calib.)	$\pm$0.8°C $-$55°C to 125°C
Trimmed Inaccuracy one-point ($\pm 3\sigma$)	$\pm$0.25°C $-$55°C to 125°C
Supply voltage	2.5 – 5.5V
Supply current (cont. op.)	39μA

5.3 A micro-power front-end

5.3.1 Overview

In this circuit, the front-end was designed to have a low power consumption, given the accuracy requirements over the full temperature range and a maximum operating temperature. Design decisions were mostly detailed throughout Chapters 3 and 4.

5.3.2 Description

In order to have a low-power front-end, efforts were done in the sense to have a small total number of current sources used to bias the PNP transistors, while having a unit current bias value set by accuracy constraints.

The front-end core and pre-bias circuit were designed with $(1+m) = (1+p) = 6$ current sources. This is sufficient to establish a 1:5 current ratio, when generating ΔV_{BE}, in both circuits, and sufficient to bias the PNP transistors when generating V_{BE}, with appropriate margin for trimming, without lack of accuracy. This was discussed with detail in Chapter 3.

By designing a low-power OTA in the pre-bias, a front-end with a total supply current smaller than 5μA was possible.

5.3.3 Experimental results

After fabrication, 15 samples from one batch were measured and characterized. Table 5.3 summarizes the main results obtained.

5.3.3.1 Batch calibration

Figure 5.6 shows the measured temperature error (PTAT frequency with $f_{nom} = 11\text{kHz}$ and third-order curvature compensation) after batch calibration (untrimmed error) with a conversion time of 2048, 1024 and 512 $\Delta\Sigma$ cycles[1], for 15 samples along with the average error and $\pm 3\sigma$ limits. Measurements show an inaccuracy below $\pm 0.8°\text{C}$ from $-55°\text{C}$ to $125°\text{C}$. This overall performance (set by the worst case error at $125°\text{C}$) is regardless of the conversion time, if 512, 1024 or 2048 $\Delta\Sigma$ cycles, and presents for all cases roughly the same dominant (PTAT) behavior[2], as depicted. However, at the low-end of the temperature range, slightly different performances can be observed (clearly more visible after trimming, as discussed as follows). In spite of the higher batch calibrated inaccuracy, trimmed inaccuracies not smaller than $\pm 0.3°\text{C}$ are expected, mainly due to the non-PTAT error components present.

5.3.3.2 Trimmed results

Figure 5.7 shows the temperature error after trimming (PTAT frequency with $f_{nom} = 11\text{kHz}$ and third-order curvature compensation). The results are comparable in terms of accuracy. However, the error shaping (butterfly shaped) is different and this is due to the available converter resolution. With 2048 $\Delta\Sigma$ cycles per conversion, the A/D conversion resolution is higher (at the price of a longer conversion time obviously), so a better trimming is achieved at the calibration temperature, i.e. at the calibration temperature the error is strongly reduced (refer to Fig. 5.7(a)) after trimming, increasing the error explosion mainly at the high-end of the temperature range. This is not observed for a conversion time of 512 cycles when the trimming performance is affected by the lack of resolution, which is the case shown in Fig. 5.7(c). In this case, the smaller performance at the calibration temperature causes the temperature error to have a more uniform behavior over temperature, which might be interesting in some applications when compared with a very small inaccuracy only at the trimming point. Additionally, at the low-end of the temperature range, the lower performance of a conversion with 2048 $\Delta\Sigma$ cycles is explained by its initial (untrimmed) poorer inaccuracy and also possibly by the relatively lower effectiveness of V_{BE} trimming at low temperatures.

[1] Different conversion times were carried out in order to better investigate issues on resolution and accuracy characteristic over temperature.

[2] While the PTAT behavior is dominant, non-PTAT errors are present.

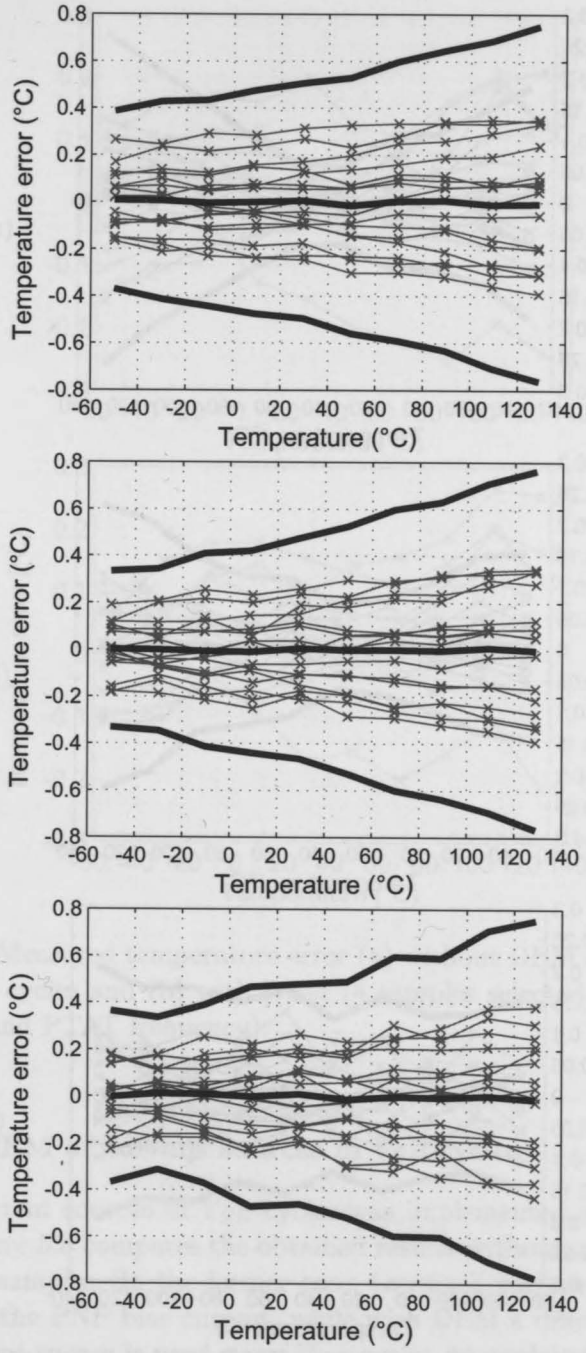


Figure 5.6: Measured temperature error after batch calibration (untrimmed) with a conversion time of (a) 2048 (b) 1024 and (c) 512 $\Delta\Sigma$ cycles (15 samples with average and $\pm 3\sigma$ limits).

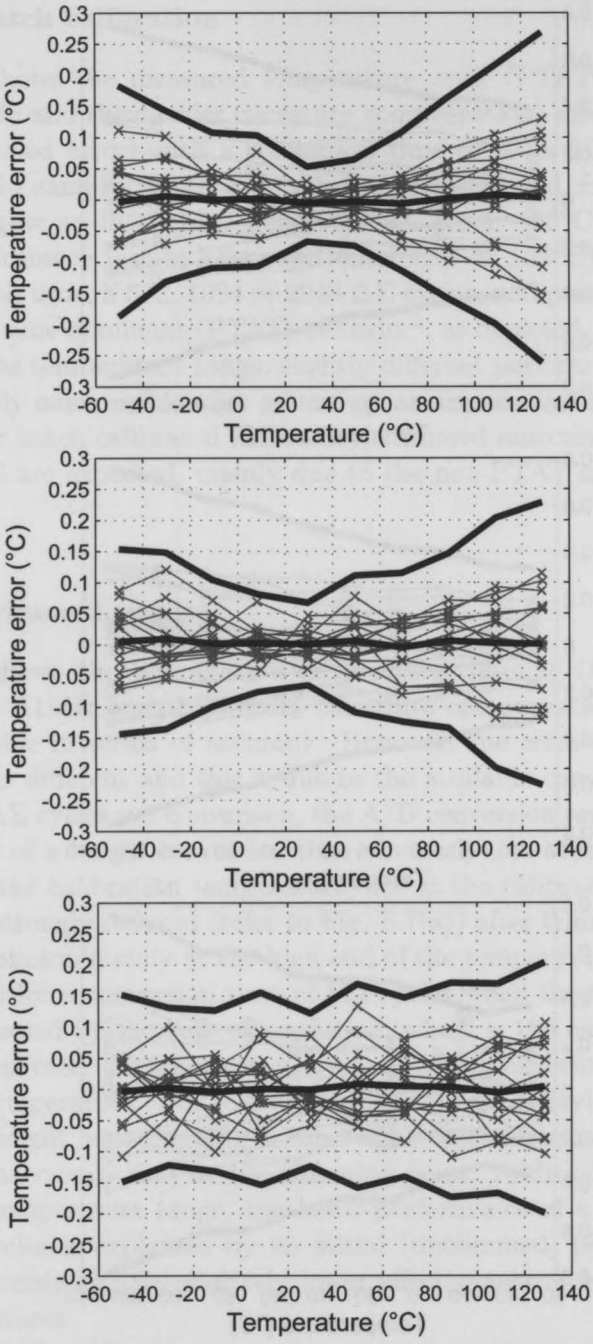


Figure 5.7: Measured temperature error after trimming at 30°C with a conversion time of (a) 2048 (b) 1024 and (c) 512 $\Delta\Sigma$ cycles (20 samples with average and $\pm 3\sigma$ limits).

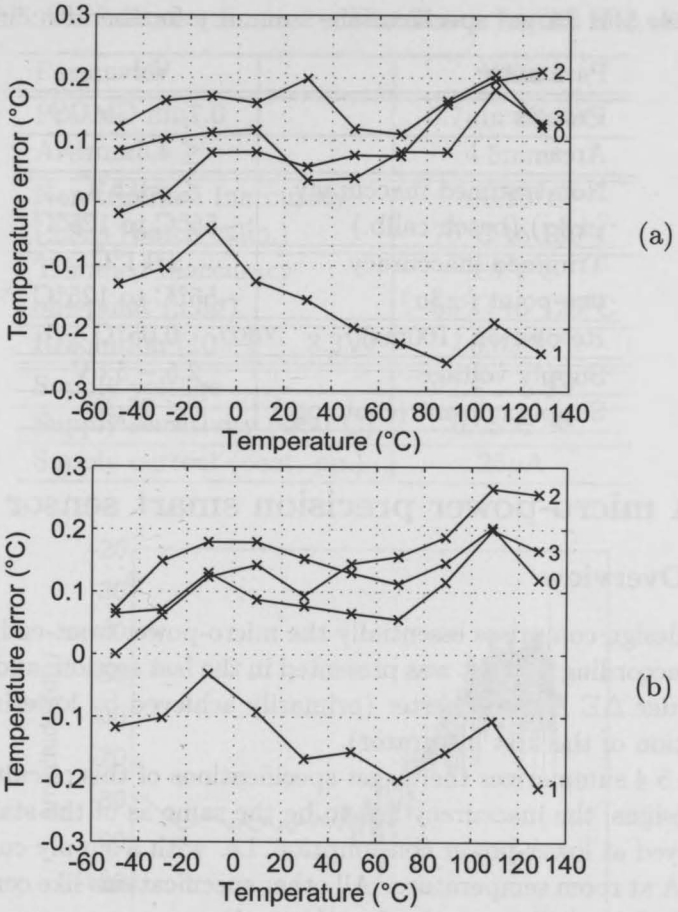


Figure 5.8: Measured temperature error (a) without DEM of current sources during V_{BE} cycles and (b) with DEM (4 samples selected at random, 1024 $\Delta\Sigma$ cycles and PTAT frequency).

5.3.3.3 DEM of current sources in V_{BE} cycles

DEM of current sources in V_{BE} cycles was implemented and tested in this design. Figure 5.8 compares the obtained results without and with DEM for 4 arbitrary samples. In the former case, two fixed current sources are used to generate the PNP bias current, while with DEM a different combination of two current source is used every V_{BE} cycle. As explained, the results are comparable, with a slightly better performance when DEM is active. This is expected, since it is the PNP saturation current I_S the dominant source of V_{BE} spread.

Table 5.4: Target specifications summary for the fifth circuit.

Parameter	Value
Process	0.7μm CMOS
Area	$\leq 4.5\text{mm}^2$
Non-trimmed inaccuracy ($\pm 3\sigma$) (batch calib.)	$\leq \pm 0.5$°C -55°C to 125°C
Trimmed inaccuracy one-point ($\pm 3\sigma$)	± 0.1°C -55°C to 125°C
Resolution (100ms)	0.05°C
Supply voltage	$2.5 - 5.5$V
Supply current (cont. op.)	$\leq 25\mu$A

5.4 A micro-power precision smart sensor

5.4.1 Overview

The last design comprises essentially the micro-power front-end, successfully designed according to what was presented in the last section, and a low-power second-order $\Delta\Sigma$ A/D converter (primarily achieved by lowering the power consumption of the first integrator).

Table 5.4 summarizes the target specifications of this circuit. Like in the former designs, the inaccuracy has to be the same as of the state-of-the-art, but achieved at lower-power consumption, i.e. with a supply current smaller than 25μA at room temperature. All other specifications like conversion time (100ms) and resolution remained unchanged.

5.4.2 Description

In this design, two major modifications were done. The first was the layout improvement of the current sources used to bias the PNPs in the front-end core. With a totally different layout and the current sources of the core and pre-bias circuitry in common centroid, matching levels were strongly improved, as well mirroring characteristics over temperature. This is, according to the author's belief, the most fundamental modification that allowed to reduce V_{BE} spread so a small batch calibrated inaccuracy was achieved compared with the former design (refer to the measurement results). Details about the current sources layout are given in Appendix H.

The second modification was the design of a high-gain low-power amplifier to implement the first integrator of the modulator. This amplifier was already detailed in Chapter 4.

Table 5.5: Performance (measured) summary for the fifth circuit.

Parameter	Value
Process	0.7μm CMOS
Area	4.5mm^2
Non-trimmed Inaccuracy ($\pm 3\sigma$) (batch calib.)	$\pm$0.25°C −70°C to 130°C
Trimmed Inaccuracy one-point ($\pm 3\sigma$)	$\pm$0.1°C −55°C to 125°C
Resolution ($10\frac{\text{conv.}}{\text{s}}$, 3.3V)	0.03°C
Supply voltage	2.5 – 5.5V
Supply sensitivity (@30°C)	0.05°C/V
Supply current (cont. op.)	25μA

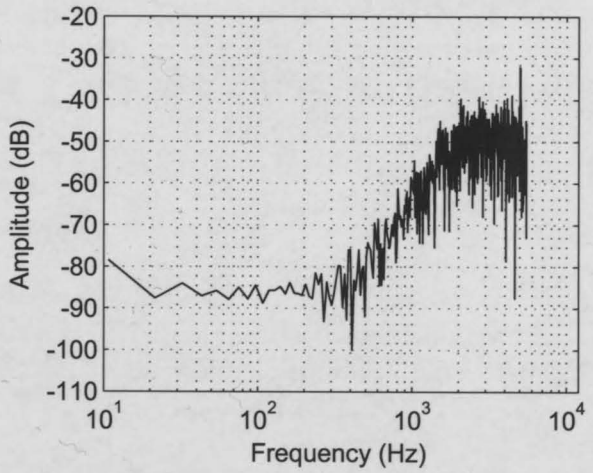


Figure 5.9: Measured power spectrum of the bitstream for a random sample (1024 bits, 4 times averaged, no windowing, PTAT frequency with f_{nom} = 11kHz).

5.4.3 Experimental results

After fabrication, 20 samples from one batch were measured and characterized. Table 5.5 summarizes the main results obtained. Figure 5.9 shows the measured bitstream power spectrum for a random sample, with no windowing, 4 times averaged, PTAT frequency with nominal frequency f_{nom} = 11kHz at room temperature.

5.4.3.1 Initial characterization

An initial characterization of some sensors are needed to perform the batch calibration[3]. The results of the characterization of these samples may then be extended for all devices. Fig. 5.10(a) presents the nonlinear temperature error for 4 samples, after linear fitting (using the least mean square approach), for all five coarse trimming modes. The linear fitting results are used to derive the linear and scalar coefficients, A and B, respectively, so the bitstream average μ can be translated to a digital temperature reading as given in (3.9), i.e. $T_{dig} = A\mu + B$. For the second coarse trimming mode, the coefficients found were $A = 578.68$ and $B = -278.56$.

The results are in good agreement with the simulation results presented in Fig. 3.24 of Chapter 2, showing the second mode as the one having the smallest nonlinear error, as designed. Fig. 5.10(b) shows this mode in detail, where a remaining third-order curvature of about $\pm 0.1°C$ is clearly visible[4]. At the high end of the temperature range, leakage introduces a considerable deviation from the ideal shape, even under a PTAT frequency scheme, luckily without increasing the error, since its effect is in opposition to the effect of the remaining third-order curvature, for this particular trimming mode.

5.4.3.2 Batch calibration results

Figure 5.11 shows the measured temperature error (PTAT frequency with nominal frequency $f_{nom} = 11$kHz and third-order curvature compensation) after batch calibration (untrimmed error), along with the average error and $\pm 3\sigma$ limits. Measurements show an inaccuracy below $\pm 0.25°C$ from $-70°C$ to $130°C$ with $V_{DD} = 2.5$-3.3V ($\pm 0.3°C$ with $V_{DD} = 5.5$V). This performance (set by the worst case error at $130°C$) is regardless of the conversion time, if with 512, 1024 or 2048 cycles ($V_{DD} = 3.3$V). However, the error behavior is different, varying from a closely temperature independent (TI) to a PTAT behavior, as compared in Fig. 5.11(a), (b) and (c). One possible explanation is due to an increased error at low temperature due to charge injection. A longer conversion might allow errors due to charge injection to accumulate during more time. Because the PTAT voltage is small at low temperature, this type of error is likely to cause more problems at the low-end of the temperature range (what it not observed at the high-end temperature range). Another explanation may be related to the system level chopping, which might have less performance in a longer conversion, particularly at low

[3]While this initial characterization is necessary (and it was performed) in all former designs, it is only presented (and detailed) here for the last design.

[4]For other modes, the second-order non linearity can add up to $\pm 0.4°C$.

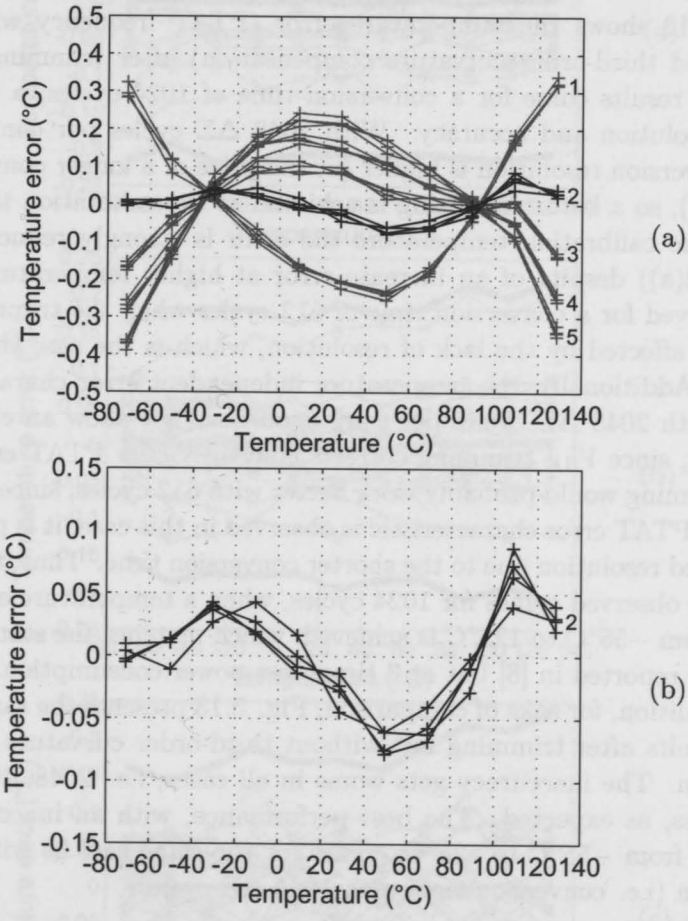


Figure 5.10: Measured non-linear temperature error (a) for all five coarse trimming modes and (b) for the second trimming mode (with the associated smallest non-linear temperature error) in detail.

temperatures. While a formal explanation was not found and it is missing, the measured error characteristic visibly changed with the number of cycles per conversion.

5.4.3.3 Trimmed results

Figure 5.12 shows the temperature error (PTAT frequency with f_{nom} = 11kHz and third-order curvature compensation) after trimming at 30°C . The best results come for a conversion time of 1024 cycles, a tradeoff between resolution and accuracy. With 2048 $\Delta\Sigma$ cycles per conversion, the A/D conversion resolution is higher (at the price of a longer conversion time obviously), so a better trimming is achieved at the calibration temperature, i.e. at the calibration temperature the error is strongly reduced (refer to Fig. 5.12(a)) despite of an increase error at higher temperatures. This is not observed for a conversion time of 512 cycles when the trimming performance is affected by the lack of resolution, which is the case shown in Fig. 5.12(c). Additionally, the temperature independent error characteristic obtained with 2048 $\Delta\Sigma$ cycles per conversion does not allow an effective V_{BE} trimming, since V_{BE} trimming corrects efficiently only PTAT errors. While V_{BE} trimming would probably work better with 512 cycles, since a more pronounced PTAT error characteristic is observed in this case, it is prevented by the limited resolution due to the shorter conversion time. Thus, the best performance observed comes for 1024 cycles, when a temperature error ± 0.1°C ($\pm 3\sigma$) from -55°C to 125°C is achieved, which matches the state-of-art performance reported in [8] but at 3 times less power consumption.

In addition, for sake of comparison, Fig. 5.13 presents the same measurement results after trimming but without third-order curvature and leakage correction. The inaccuracy gets worse in all cases, i.e. 2048, 1024 and 512 $\Delta\Sigma$ cycles, as expected. The best performance, with an inaccuracy below ± 0.15°C from -55°C to 125°C, comes for the same case as with curvature correction (i.e. conversion time with 1024 $\Delta\Sigma$ cycles).

Limiting the trimming deepness

This procedure was rapidly developed to improve the achievable trimming performance when the V_{BE} trimming method is used with sensors having a temperature independent error behavior, like that shown in Fig. 5.11(a)[5]. Performing V_{BE} trimming of such sensors is not efficient, and the error after trimming, often butterfly shaped, is still high[6]. This is mainly because en-

[5]While this procedure may need more study, valuable results were obtained.

[6]A digital offset correction surely would give better results .

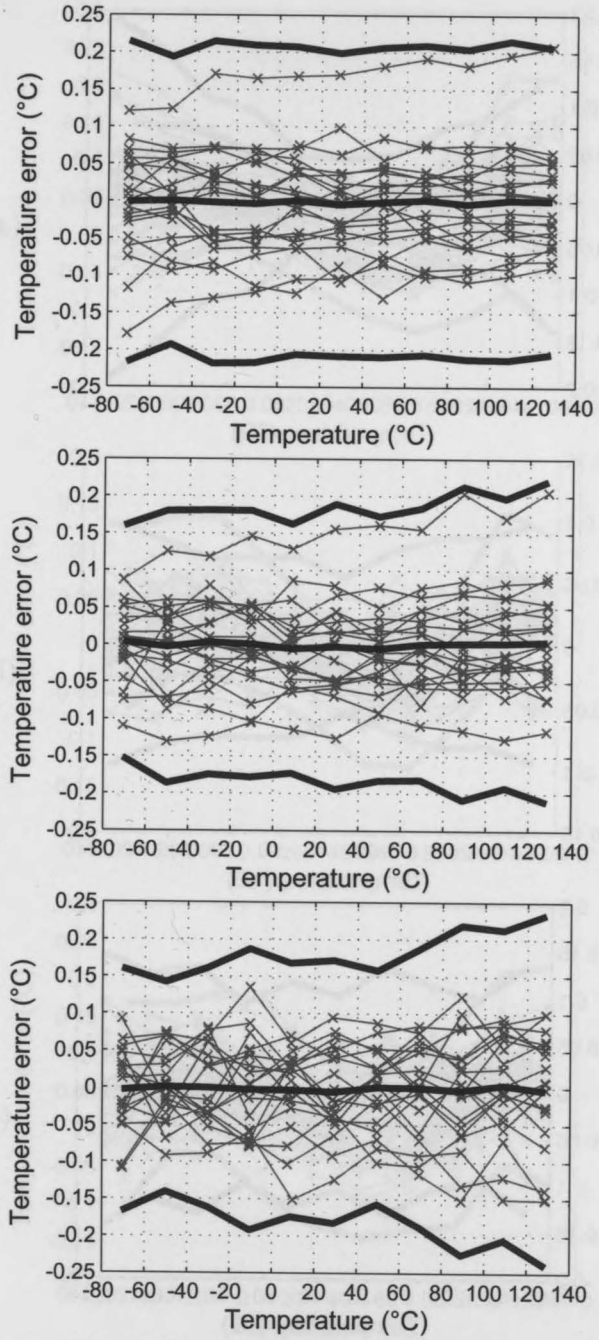


Figure 5.11: Measured temperature error after batch calibration (untrimmed), for a conversion time of (a) 2048 (b) 1024 and (c) 512 $\Delta\Sigma$ cycles (20 samples with average and $\pm 3\sigma$ limits).

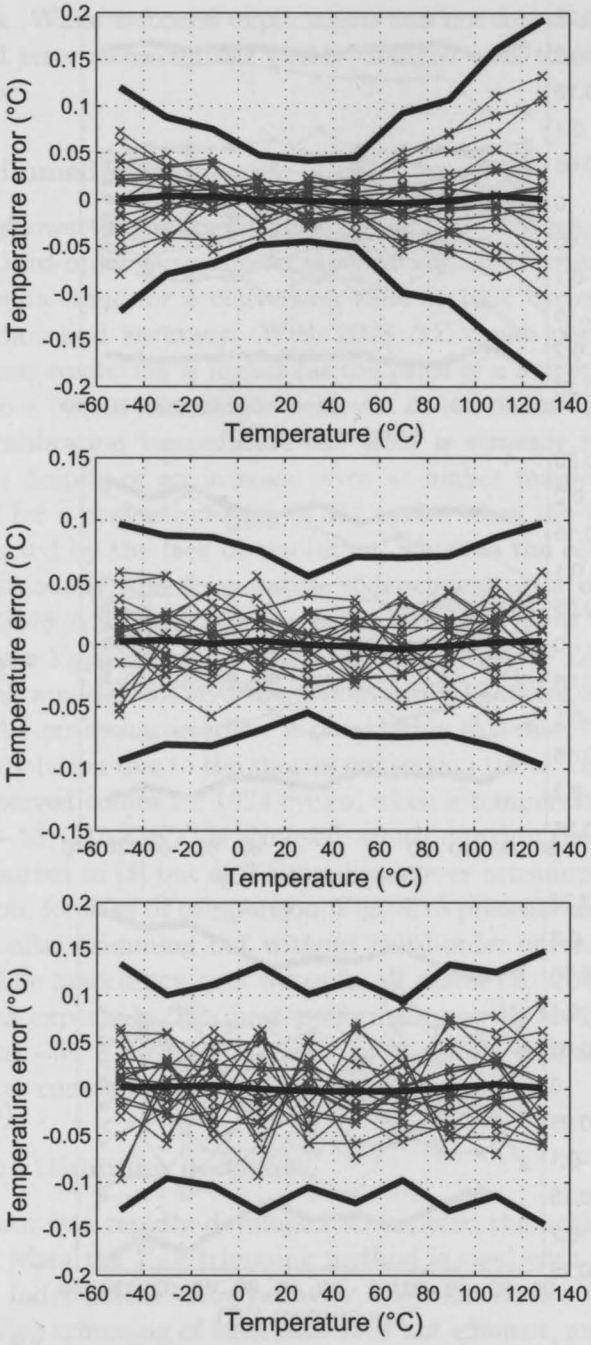


Figure 5.12: Measured temperature error after trimming at 30°C for a conversion time of (a) 2048 (b) 1024 and (c) 512 $\Delta\Sigma$ cycles (20 samples with average and $\pm 3\sigma$ limits).

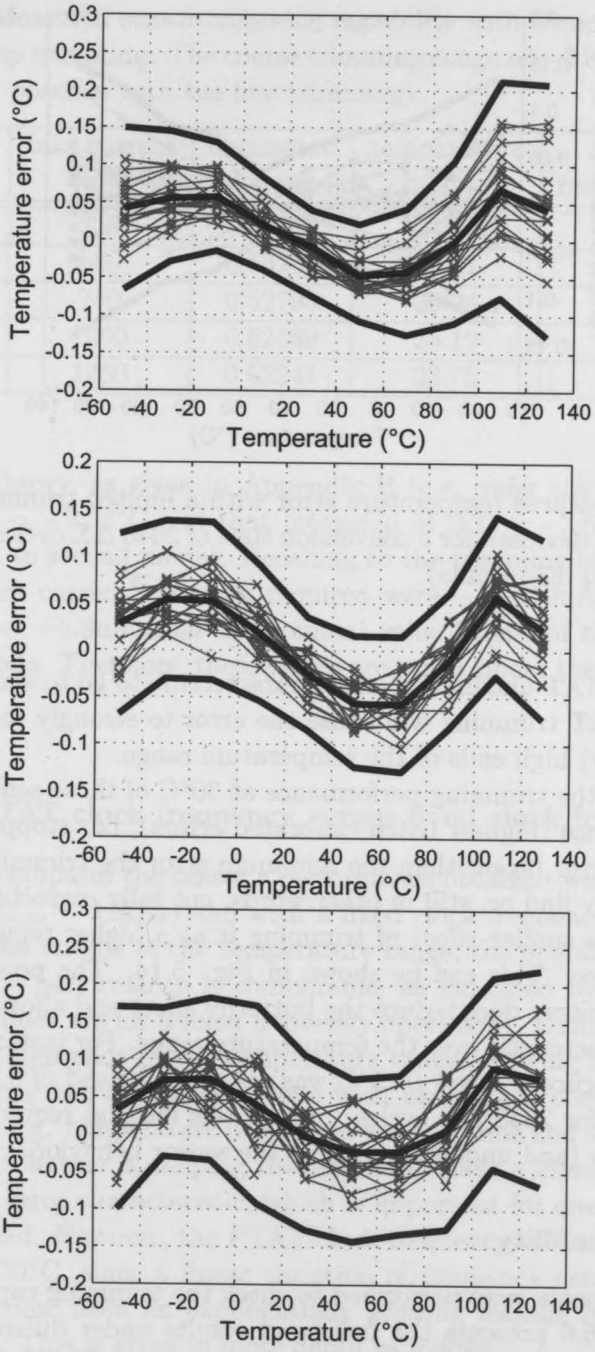


Figure 5.13: Measured temperature error (without third-order curvature compensation) after trimming at 30°C for a conversion time of (a) 2048 (b) 1024 and (c) 512 $\Delta\Sigma$ cycles (20 samples with average and $\pm 3\sigma$ limits).

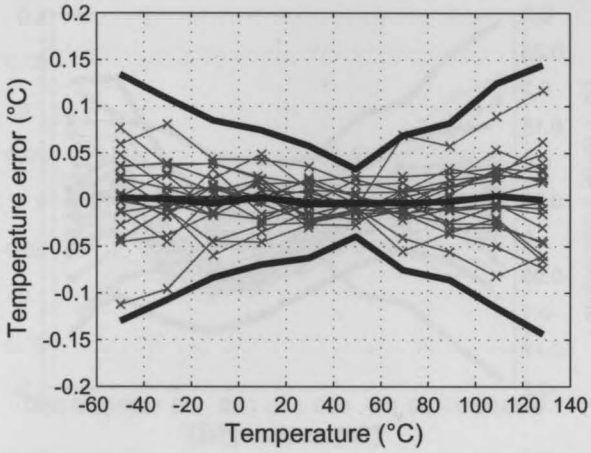


Figure 5.14: Measured temperature error with a limited trimming deepness at 30°C (4 worst devices), for a conversion time of 2048 $\Delta\Sigma$ cycles (20 samples with average and $\pm 3\sigma$ limits).

forcing a non-PTAT untrimmed error characteristic to go to zero at 30°C by means of a PTAT trimming will cause the error to strongly increase at the low and (mainly) high ends of the temperature range.

By limiting the trimming performance at 30°C of those sensors with the worst performance (highest batch calibrated errors), i.e. stopping the trimming with an error bigger than the minimum error the trimming procedure would eventually find or, still in other words, not fully correcting the sensor operation, has a similar effect of trimming it at a higher temperature, e.g. 50°C in this case. This can be shown in Fig. 5.14. The poorer trimming at room temperature does reduce the butterfly effect and allows a more uniform error characteristic over the temperature range. For instance, with this method an inaccuracy below 0.15°C was obtained instead of 0.2°C as previously reported for 2048 $\Delta\Sigma$ cycles. Surely, this method requires a previous characterization (and understanding) of the sensor behavior.

Trimming capability

Some measurements were performed to check the trimming capability of the sensor. Table 5.6 presents the measured results under different trimming conditions obtained for a random sample[7]. The results are in a good agree-

[7]Sensor #8, $A = 578.68$ and $B = -278.56$, @30°C, 29.607°C according to Pt100 resistors, 512 $\Delta\Sigma$ cycles.

Table 5.6: Measured sensor trimming capability with 5 current sources used as coarse-step trimming. The coarse trimming range is [−5.86,+4.49] (a small extension is possible with the fine trimming).

trimming mode #	bias current (nA)	decimated bitstream	temperature (°C)	ΔT(°C) with respect to mode 2
1	250	0.54029	34.10	+4.49
2	500	0.53252	29.61	0
3	750	0.52793	26.94	−2.67
4	1000	0.52489	25.19	−4.42
5	1250	0.52241	23.75	−5.86

ment with theory, as given in Appendix B (e.g. refer the value calculated there of $\partial T = -2.58$°C and that measured of $\Delta T = -2.67$°C of the third with respect to second mode). Recalling to the trimming of the 20 samples, the maximum current variation required were $-\frac{13}{255}250\text{nA} \approx -12.7\text{nA}$ and $+\frac{9}{255}250\text{nA} \approx +8.8\text{nA}$ from the nominal value 500nA of the second coarse trimming mode. Therefore, these values are surely under the sensor trimming capability.

5.4.3.4 PTAT clock frequency versus fixed clock frequency

Figure 5.15 compares the sensor's performance obtained with a PTAT clock frequency ($f_{nom} = 11\text{kHz}$) and with a fixed clock frequency ($f = 11\text{kHz}$)[8]. Because in the middle of the temperature range, the operating frequency is the same, the performance is comparable, as indicated by the $\pm 3\sigma$ limits (the dashed lines are for a fixed frequency operation). Differences appear at the low and high ends of the temperature range, showing how an adaptive clock frequency can successfully cope with issues that come with a PTAT bias current, e.g. temperature dependent settling time and signal leakage, as explained in Chapter 4, helping the sensors to have a more PTAT behaved temperature error characteristic, which is important for one-point trimming for V_{BE} spread. However, the PTAT clock frequency technique works properly up to 130°C, since a linear variation of frequency cannot compensate errors that come from an exponentially growing leakage further then this temperature. This is given in more detail as follows.

[8]Because the batch calibration results differ in both cases, different scaling factors A and B were used so to properly compare the two cases.

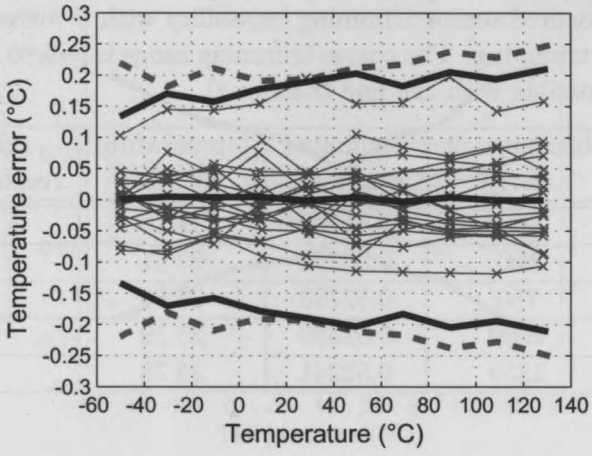


Figure 5.15: Measured temperature error using a PTAT clock frequency (with a nominal frequency $f_{nom} = 11\text{kHz}$) and fixed clock frequency ($f = 11\text{kHz}$) (20 samples with average and $\pm 3\sigma$ limits). Dashed lines are for a fixed frequency operation.

5.4.3.5 Measurement at 150°C

The 20 samples were tested up to 150°C. At these temperatures, however, the PTAT clock cannot completely compensate for the exponential rise in leakage current[9]. Frequencies higher than that defined by a PTAT characteristic are needed. Table 5.7 shows some frequencies used at 150°C. The measured results show a systematic non-linearity varying accordingly to the clock frequency used. For instance, at 150°C, a systematic error (in average) of −0.9°C, −0.5°C and 0°C (± 0.25°C, $\pm 3\sigma$ limits are not shown) is observed for clock frequencies of 20kHz, 25kHz and 35kHz, respectively. This is graphically shown in Fig. 5.16. Because of the nature of such error that presents a very systematic behavior like the third-order curvature also has (indeed very visible in the Fig. 5.16), the same non-linear filtering used to handle the curvature can be used to compensate for leakage at 150°C, which might extend the sensor's operating range with inaccuracies below ± 0.25°C, as it is the case for temperatures up to 130°C.

[9]A nonlinear frequency oscillator is needed to suppress the effect of leakage at high temperature, which implementation might be complex.

Table 5.7: PTAT clock frequency - System operating clock as a function of temperature (or front-end bias current). At 150°C, a 35kHz is needed to prevent leakage errors of introducing large temperature errors.

temp. (°C)	from -50 to 130	150
clock frequency (kHz)	7.5-17.5 (PTAT)	20, 25, 35

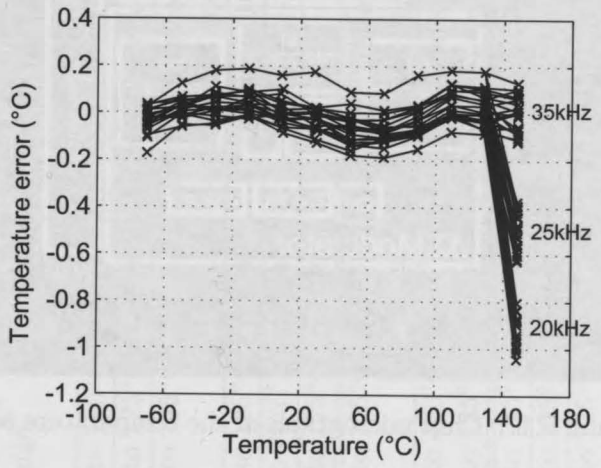


Figure 5.16: Measured temperature error up to 150°C with three different frequencies: 20kHz, 25kHz and 35kHz. (20 samples with no third-order curvature correction).

5.4.3.6 Chip micrograph

A micrograph of the temperature sensor is shown in Fig. 5.17 where some main building blocks are detailed. The chip does not include the decimation filter, some digital circuitry used for testing and the frequency-dependent oscillator (used to implement the PTAT frequency) in order to keep flexibility while testing the sensor, e.g. easiness while programming number of cycles, chopping frequency, coarse and fine trimming, DEM of current sources and capacitors, operation frequency, etc. The circuit has a total area of 4.5mm^2.

5.4.4 Benchmark

Table 5.8 summarizes the performance of the last two designed sensors and compares them with other academic and commercial designs (refer also to the survey [10] presented in Chapter 1). The micro-power smart sensor has a 2 times smaller batch-calibrated inaccuracy from -70°C to 130°C and

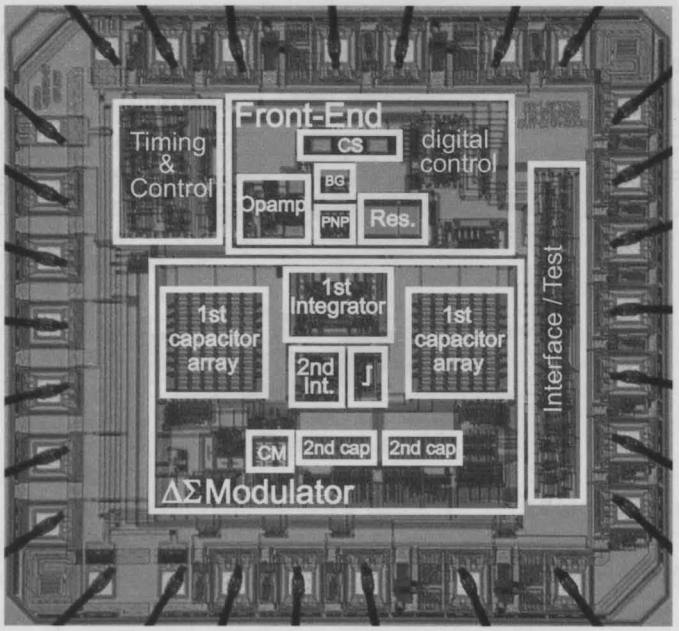


Figure 5.17: Chip micrograph of the temperature sensor.

the same trimmed inaccuracy but at 3 times less power consumption when compared to the state-of-art sensor at the time of publication[8]. Some recent works reporting designs published after this thesis project was completed were included for sake of a more updated benchmark.

Table 5.8: Inaccuracy, continuous supply current and operating range comparison of some smart temperature sensors. FOM (energy × error%^{2} in nJ%^{2}) column mostly calculated with data from [10].

Reference	Inaccuracy (°C)	Supply current (μA)	Operating range	FOM (nJ%^{2})	Remarks
Bakker [44]	±1°C (max)		−40°C to 120°C	1.7k	2-point trim. (3 samples)
Tuthill [40]	±1.5°C (max)	1k	−55°C to 125°C	100	1-point wafer level (6 samples)
Pertjis [8]	±0.5°C (±3σ)	75	−55°C to 125°C		batch calibration (24 samples)
	±0.1°C (±3σ)			231	1-point trim. (cal. with Pt100)
Vroonhoven [18]	±0.5°C (±3σ)	500(est.)	−55°C to 125°C	2300k	no trim. (16 samples)
ADT7301 [85]	±1°C (max)	> 190	0°C to 70°C	46k	-
TMP275 [58]	±1°C (max)	50	−40°C to 125°C	44k	-
This work,	±0.8°C (±3σ)	39	−55°C to 125°C		batch calibration (16 samples)
Section 5.4	±0.3°C (±3σ)			1.1k	1-point trim. (cal. with Pt100)
This work,	±0.25°C (±3σ)	25	−70°C to 130°C		batch calibration (20 samples)
Section 5.5 [9]	±0.1°C (±3σ)		−55°C to 125°C	76	1-point trim. (cal. with Pt100)
Souri [13]	±0.25°C (±3σ)	6	−40°C to 125°C	82	1-point αtrim. (19 samples)
Vroonhoven [86]	±0.2°C (±3σ)	600(est.)	−55°C to 125°C	926k	no trim. (16 samples)
Sebastiano[12]	±0.5°C (±3σ)	8.3	−70°C to 125°C		untrimmed (16 samples)
	±0.2°C (±3σ)			190	1-point trim.
Souri	±0.25C (±3σ)	4.6	−30°C to 125°C		batch calibration (19 samples)
	±0.2°C (±3σ)			49	1-point trim.

5.5 Conclusions

In this chapter, measurements results of five designs of a CMOS smart temperature sensor were presented. Although having a similar core, each one has its own characteristics and thus contribution. In this sense, despite of the former sensors having somewhat a low performance, each one was an effective step towards the last sensor, remarkable for its non trimmed inaccuracy below $\pm 0.25°C$ over an extended military temperature range, i.e. $200°C$ from $-70°C$ to $130°C$. Since trimming is a major cost contribution of a sensor, the achievement of such untrimmed low inaccuracy with a total supply current of $25\mu A$ is certainly a major achievement, enabling this sensor for a great number of applications. This is reinforced by the benchmark presented at the end of this chapter.

CHAPTER 6

Conclusions

In this final chapter and end of work, after having worked with different variants of a temperature sensor design, and struggled against accuracy loss under power constraints, and also carried out many measurement-results analysis, the main findings are listed. In addition, original contributions are given with regard to the prior-art [22][8] and current reported literature. As a final contribution, some future work in this field is suggested.

6.1 Main findings

The design and measurement of five sensors have revealed some and corroborated other important design aspects of integrated temperature sensors. The main findings of this work are listed below:

- In spite of other temperature sensing elements (recall to Table 2.1 in Chapter 2), substrate PNP transistors (a bipolar transistor which is compatible with standard CMOS technology), are still the best choice for implementation of low-power, precision and absolute temperature sensors (Chapter 2)

  The PNP area of 10x10 μm^2 in the final sensor design is approximately 5 times smaller than the area reported in the prior-art, and the obtained inaccuracy levels are similar (trimmed results) or better (batch-calibrated results) when compared to it. This stresses the fact the PNP transistor performs very well as temperature sensing element. Besides, temperature sensors recently developed in submicron technologies using bipolar transistors [13][87][12] have also shown very

promising low-power and accurate temperature sensor operation[1];

- A non-ideal β-compensation due to a β-mismatch has the same physical effect on V_{BE} of that of β-spread: it causes V_{BE} spread. Since the β-mismatch (which is an on-chip variation) is smaller than the β spread (a chip-to-chip variation), a small contribution to V_{BE} spread is expected though (Chapter 2)

  However, while the V_{BE} spread due to the saturation current I_S is strongly process related and cannot be handled, V_{BE} spread due to a non-ideal β-compensation can, e.g. by improving PNPs β-matching levels and guaranteeing an accurate β-mirroring from the front-end pre-bias circuit to the core;

- In spite of having current ratios m and p in the core and pre-bias, respectively, independent from each other, by setting them equal improves PNPs β-matching levels in the sensor front-end (Chapter 3)

  Although the current gain β in vertical PNPs is weakly dependent on the biasing for some current decades (AMIS0.7μm [28] but also in other technologies [15][29]), a small variation is always expected for different bias. Therefore, using the same current ratio in the core and in the pre-bias circuit will help reducing errors caused by PNP β-mismatch. This is likely due to the fact that the PNP transistors in both sides are operating under more similar conditions, i.e. current ratio and bias. Also, a factor 1:5 causes the left and right transistors to operate under less bias difference when compared to ratios like 1:10, as in the prior-art and other works reported in the literature;

- A high performance (over temperature) front-end current mirror is needed to get batch-calibrated sensors with low inaccuracies (Chapter 3)

  Since current sources mismatch is temperature dependent, by better laying out the front-end current mirror, and placing the current sources of the core in between the current sources of the pre-bias (common centroid technique) helps to reduce temperature effects over the mirroring function. This reduces PTAT errors resulting from the current sources mismatch while avoid introducing other non PTAT ones, e.g. due to the β-compensation scheme. Efforts in the front-end current mirror helped significantly the achievement of batch-calibrated (without individual sensor trimming) inaccuracies below ± 0.25°C ($\pm 3\sigma$) over 200°C;

[1] In the latter, however, a bipolar NPN transistor was used instead, available in a double-well CMOS technology.

- The PTAT bias current used to set more accurately the PNP's bias currents, thus causing a PTAT spread on V_{BE} (as opposed to other bias types, e.g. TI or CTAT), but also used to diminish the V_{BE} curvature, by reducing temperature errors due to V_{BE} non-linearity over temperature, has two main drawbacks (Chapter 3 and 4):

  -The first one is the temperature-dependent CMOS current-mirrors matching levels. Since the current mirrors have in essence a constant g_m over temperature[2] (less than 2% variation from the nominal value, $\approx$ 660nA/V at 30°C, according to simulations performed at −55°C, 30°C and 125°C), a PTAT bias current causes the ratio $\frac{g_m}{I_{bias}}$ to decrease as the temperature increases (about ±50 % from the nominal value). Since matching levels are dependent of this ratio (inversely proportional to $\frac{g_m}{I_{bias}}$) the consequence is a current mirror performance that is affected by the temperature, providing a poorer (or better) matching at lower (or higher) temperature.

  -The second one is the temperature-dependent settling time, since the front-end sensor has a driving capability, due to the PTAT current, dependent on the temperature. Often the common approach or solution is to design the system clock for the worst case, i.e. lowest temperature, but this is undesired in case of stringent power constraints, since it means waste of power at higher temperatures; the inverse is lack of time for settling if adequate time is set at high temperature (Chapter 4). An adaptive frequency (PTAT frequency) was proposed as a solution for the later case, while the worst case scenario adopted for the former, despite of layout improvements as discussed;

- PNPs bias reduction and A/D converter's sampling-capacitors scaling down were shown to enhance circuit non-idealities which might degrade sensor's performance when not canceled or compensated by some circuit technique (Chapter 3 and 4)

  The PNP saturation current (estimated to a hundred of pA at higher temperatures), in the presence of a low bias current (some hundreds of nA), act degrading significantly the current ratio m and hence an accurate PTAT voltage generation and ratiometric temperature measurement. Additionally, the A/D converter's capacitors scaling down (so to keep the sensor's bandwidth unchanged given the lower bias current) may degrade the sensor's accuracy by increasing the sensitivity of

[2]The variation of g_m due to a PTAT bias current is compensated by the variation of the mobility μ, that decreases with the temperature.

the A/D conversion to charge injection and signal leakage effects. Advanced instrumentation techniques, though, effectively work with the main drawbacks resulting of the smaller bias currents and capacitors. Without such techniques, however, it would be difficult to obtain temperature errors of 0.1°C over the military range (while having a power consumption below a hundred of μW).

- The modified (but more complex) voltage V_{TEMP}, thought for a more efficient use of the available A/D converter input range, while implementable, is in practice not effective to reduce power consumption (Chapter 4)

  While the voltage V_{TEMP} uses more efficiently the available A/D converter input range, its longer $\Delta\Sigma$ cycle cancels the benefit of the smaller needed conversion resolution (and hence number of $\Delta\Sigma$ cycles), since in the best scenario, the length of the original $\Delta\Sigma$ cycle is doubled, given the target architecture addressed in this work. In addition, errors due to charge injection are prone to increase because of the more complex $\Delta\Sigma$ cycle the voltage V_{TEMP} needs;

- $\Delta\Sigma$ A/D converters are the most suitable data converters for accurate smart temperature sensors, in addition to their ability to operate at low-power (Chapter 4)

  This finding reaffirms what was stated in [22] while it makes it more general, as it takes into consideration not only the achievable performance in terms of accuracy, but also in terms of accuracy under power constraints. However, it does not stand if a low- (or ultra-low) power characteristic is the most important specification of a sensor, in which case other architectures like SAR (Successive Approximation Register) or Algorithmic will result in less energy per conversion. Furthermore, mixed approaches like Extended-Counting [62] or Zoom A/D converters [13] will also result in more energy-efficient solutions, though still relying on $\Delta\Sigma$ architectures;

- With respect to the modulator, there is still margin for third-order architectures, since the input signal ΔV_{BE} hardly uses $\frac{1}{3}$ of the available input range of the A/D converter. In addition, a discrete-time implementation fits better the particular needs of the application (Chapter 4)

  In this sense, a more aggressive noise shaping with a third-order modulator is possible, having in theory advantage in terms of energy per conversion with no issues on stability due to the higher filter order.

Besides, because discrete-time (switched-capacitors) modulators can work more accurately with crude clock oscillators (when compared to continuous-time modulators), the proposed PTAT frequency technique may have a simpler implementation in the former case.

6.2 Original contributions

In order to design a precision low-power temperature sensor, error sources and accuracy trade-offs needed to be systematically addressed, either by circuit or system level techniques, in all design steps. Regardless of the step, original design approaches combined with prior knowledge and optimization resulted in some contributions, that to the best knowledge of this author, are original. They are:

- Establishment of some critical accuracy-power consumption trade-offs present in the design: front-end and $\Delta\Sigma$ A/D converter, which allowed to set small and close-to-minimum bias currents in this circuit given accuracy requirements (Chapter 3 and 4)

  A minimum PNP bias current is set given a maximum temperature error contribution to the final accuracy. In addition to current amplitudes, current ratios, both in the front-end core and pre-bias circuit, were also derived. Particularly in the pre-bias, a current ratio of 5, equal to the current ratio in the core, was found more appropriate in terms of accuracy and power consumption. In the $\Delta\Sigma$ A/D converter, a minimum OTA tail current (first integrator) was set given a minimum sampling capacitor set by accuracy requirements;

- Mismatching analysis over temperature of current sources in the front-end (Chapter 3)

  Because of the PTAT current used to bias the PNP transistors, mismatching levels of CMOS current sources are temperature dependent and so does the mirroring and the β-compensation scheme performance. Due to the importance of the mirroring in the front-end, simulations were performed to better understand and characterize this function in the front-end;

- Balanced trimming (Chapter 3)

  Based on the prior knowledge, a more balanced trimming range is set if two (of five) unit current are used to bias the PNP transistors. Although the trimming resolution is unbalanced due to the logarithmic

dependence of V_{BE} on the bias current I_{bias}, a balanced trimming capability (of about $\pm 5°C$) is available;

- Dynamic Element Matching of current sources in V_{BE} cycles (Chapter 3)

  In addition to DEM of current sources in ΔV_{BE} cycles, already reported, DEM of current sources in V_{BE} cycles was introduced. Two current sources (of five) are used in a V_{BE} cycle. By using in every V_{BE} cycle a different combination of two unit current sources, an effective smaller bias mismatch is achieved when generating V_{BE}, which helps improving the sensor's batch calibrated performance, which is mainly and ultimately set by the PNP's saturation current spread;

- PTAT clock frequency (Chapter 4)

  A temperature dependent (PTAT) operating clock frequency was proposed to avoid the drawbacks of the PTAT current used to bias the PNP transistors. The PTAT clock frequency can be implemented with a temperature dependent oscillator or with a Voltage Controlled Oscillator (VCO) whose control may be derived from the PTAT voltage ΔV_{BE}. By modifying the clock frequency accordingly to the temperature, an improved operation performance over the temperature range is verified given tight accuracy and power constraints;

- Temperature error characterization versus the number of $\Delta\Sigma$ cycles per conversion (Chapter 5)

  Measurement results obtained for different conversion times, i.e. number of $\Delta\Sigma$ cycles per conversion (given thus the same nominal operating frequency) have shown different temperature error characteristics and the trade-off resolution and accuracy;

- Trimming deepness (Chapter 5)

  Limiting the sensor trimming (correction) at the calibration temperature may improve the sensor's overall performance (over temperature) when a PTAT trimming method (V_{BE} trimming) is used to correct a non PTAT, e.g. TI temperature error characteristic. Although not a formal method (it is much more a way to cope with the limitations of V_{BE} trimming in face of TI temperature errors), a superior performance over temperature was obtained, despite of an expected larger error at the calibration point. A prior knowledge of the sensor's characteristics is needed though;

- Establishment of a new state-of-the-art batch-calibrated (non-trimmed) inaccuracy, i.e. $\pm 0.25°C$ ($\pm 3\sigma$) over a wider military temperature range, i.e. $-70°C$ to $130°C$ (Chapter 5)

  Since most of the error sources were canceled by design (following the knowledge developed in the prior art), the remaining main error is that due to V_{BE} spread. Specific efforts to reduce this spread were carried out, e.g. current mirrors improved layout (more compact and nested common-centroid), attenuating the temperature dependence in the mirroring function as well the β-compensation scheme. In addition, using the same current ratios in the core and pre-bias circuits might have contributed to reduce spread.

6.3 Future work and perspectives

It is not seldom, at the end of a work, to have some implementation ideas left over. They are thus mentioned here as future work. Perspectives, unlikely, are more a projection, having always a degree of guessing. The reader is asked to take this into account.

6.3.1 Future work

Some suggestions for future work are given below:

- Ultra-low power smart temperature sensor design

  Given the trend for autonomous systems [16] and sensor networks [5][88], it is important to think of ultra-low power temperature sensors (without a major concern for high accuracy). In such energy- and power-restricted systems, the ability to run and provide a temperature reading under less strict accuracy requirements, however, dissipating continuously , e.g. $1\mu W$ or less, is certainly welcome.

- On-chip PTAT-frequency oscillator design

  In order to implement the PTAT clock frequency technique needed to counter-balance the drawbacks of the PTAT current used to bias the PNP transistors in the front-end, an on-chip PTAT-frequency oscillator is required. Such oscillator can be implemented with a VCO, widely applicable and a fundamental building block in many systems [89]. Alternatively, a relaxation oscillator with a temperature dependent biasing can produce the needed PTAT output clock frequency. Regardless of the case, the power dissipated by the oscillator should add a small fraction of the total temperature sensor power consumption.

- Decimation filter design

  Whether an on-chip decimation filter is more advantageous or not (since an off-chip filter is more flexible and can be tuned to a particular sensor application), it is convenient to have it internally available. A mechanism though to bypass it, in the presence of an external filter, has to be included. A more specific and careful work about low-power decimation filters for smart sensors is missing.

- Third-order CIFF-IF $\Delta\Sigma$ design

  A third-order incremental $\Delta\Sigma$ A/D converter, instead of a second-order one, might result in a smart temperature sensor with a smaller power consumption and same accuracy levels reported here. In addition, the combination of $\Delta\Sigma$ architectures with other architectures, e.g two-step converters [60], Extended-Counting [62][90] and Zoom converters [13] might result in even less power consumption, solved standing low-accuracy issues.

6.3.2 Perspectives

Sensors of all colors and flavors for measuring a large variety of physical quantities are widely available and already in use in many applications. And there is no doubt the demand for sensors will continue increasing for many years. The question is then not if sensors will be needed, but how to provide the accuracy required by many applications, with the power consumption, bandwidth, cost, etc. required by many others. The answer seems to be in the correct and efficient exploitation of circuit and system level techniques to cope with the sensor's electronics non-idealities.

This work is one example where high performance was achieved by a systematic error cancellation, in addition to an optimization of the critical accuracy-power trade-offs present in the design. For instance, various mismatch-related errors were eliminated through the use of precision layout, dynamic offset cancellation and dynamic element matching (DEM). Furthermore, errors caused by incomplete settling and leakage were reduced by the use of a temperature-dependent clocking scheme (PTAT clock frequency). The less amount of trimming was possible by keeping the V_{BE} spread to minimum values set by the fabrication process technology.

Of course, newer CMOS technologies may provide more performance, by offering higher-quality lithography or by means of a more controlled fabrication steps (with reduced spread of process parameters). But newer technologies are in general more expensive while typically sensors have to be low cost, usually a tiny fraction of the total system price. This re-states the

need for the development of techniques capable of coping with the electronics non-idealities, so sensors will continue meeting the applications needs.

APPENDIX **A**

PNP transistors-mismatch effect on ΔV_{BE}

The substrate PNP transistor (in the front-end) biased through the emitter with a current I_{bias} provides a voltage V_{BE} given by

$$V_{BE}(T) = V_T \ln\left(\frac{I_C}{I_S(T)} + 1\right) \approx V_T \ln\left(\frac{I_{bias}}{I_S(T)}\right) \tag{A.1}$$

where $V_T = \frac{kT}{q}$ is the thermal voltage and I_S is the saturation current.

A voltage $\Delta V_{BE} = V_{BE,2} - V_{BE,1}$ proportional to absolute temperature (PTAT) is obtained by biasing two PNPs Q_1 and Q_2 at $1 : m$ current ratio, i.e. $m = \frac{I_{bias,2}}{I_{bias,1}} = \frac{(m+\Delta m)I_{bias,1}}{I_{bias,1}}$, where m is the error in the current ratio due to bias current mismatch. Provided errors due to mismatch in the bias currents ($\Delta m = 0$) and PNP transistors are negligible ($I_{S,1} = I_{S,2}$) , this voltage is given by

$$\Delta V_{BE}(T) = V_T \ln(m). \tag{A.2}$$

However, if PNP transistors are not matched as well the bias currents, this voltage is more accurately written as

$$\begin{aligned} \Delta V_{BE,21}(T) &= V_{BE,2} - V_{BE,1} \\ &= V_T \ln\left(\frac{\frac{I_{bias,2}}{I_{S,2}}}{\frac{I_{bias,1}}{I_{S,1}}}\right) = V_T \ln\left(\frac{(m+\Delta m)I_{bias}}{I_{bias}}\frac{I_{S,1}}{I_{S,2}}\right). \end{aligned} \tag{A.3}$$

The error in the current ratio can be solved (or strongly reduced) with DEM, when $\Delta V_{BE,21}$ is given by

$$\Delta V_{BE,21}(T) = V_T \ln\left(\frac{mI_{bias}}{I_{bias}}\frac{I_{S,1}}{I_{S,2}}\right). \tag{A.4}$$

But the error due to the PNP transistors mismatch cannot. One solution is to switch the currents $I_{bias,1}$ and $I_{bias,2}$, biasing Q_1 and Q_2 at $m:1$ current ratio, establishing a second $\Delta V_{BE,12}$ given by

$$\Delta V_{BE,12}(T) = V_{BE,1} - V_{BE,2} \tag{A.5}$$

$$= V_T \ln\left(\frac{\frac{I_{bias,1}}{I_{S,1}}}{\frac{I_{bias,2}}{I_{S,2}}}\right) = V_T \ln\left(\frac{mI_{bias}}{I_{bias}}\frac{I_{S,2}}{I_{S,1}}\right). \tag{A.6}$$

An averaged ΔV_{BE} can be computed by adding together $\Delta V_{BE,21}$ and $\Delta V_{BE,12}$

$$\Delta V_{BE,21} + \Delta V_{BE,12} = V_T \ln\left(\frac{mI_{bias,1}}{I_{S,2}} \cdot \frac{I_{S,1}}{I_{bias,1}}\right) + V_T \ln\left(\frac{mI_{bias,1}}{I_{S,1}} \cdot \frac{I_{S,2}}{I_{bias,1}}\right)$$

$$= 2V_T \ln(m) \tag{A.7}$$

and dividing it by two, i.e.

$$\Delta V_{BE,avg} = V_T \ln(m) \tag{A.8}$$

This result shows that the effect of the PNP transistor mismatch was canceled.

APPENDIX B

Trimming capability

The sensitivity of T with respect to V_{BE} is given as

$$\frac{\partial T}{\partial V_{BE}} = -\frac{V_{REF}}{T} \approx \frac{0.25°\text{C}}{\text{mV}} @ T = 300\text{K},\ V_{REF} = 1.2\text{V}. \tag{B.1}$$

For a trimming capability of $[4.49, -5.86]°\text{C}$, as needed and measured (refer to Table 5.6 of Chapter 5), i.e. $\Delta T_{total} \approx 10.35°\text{C}$ or $\pm 5.17°\text{C}$ (if the range is centered), a total trimming of $4\text{mV}/°\text{C} \times \pm 5.17°\text{C} = \pm 20.68\text{mV}$ ($|\Delta V_{BE}| \approx 41.36\text{mV}$) is needed for V_{BE}.

Since the voltage V_{BE} is given as

$$V_{BE} = V_T \ln\left(\frac{I_C}{I_S}\right) \approx V_T \ln\left(\frac{I_{bias}}{I_S}\right) \tag{B.2}$$

(where the effect of the finite β is neglected), a V_{BE} variation (ΔV_{BE}) due to spread in I_S (ΔI_S) can be written as

$$\begin{aligned} V_{BE} \pm \Delta V_{BE} &= V_T \ln\left(\frac{I_{bias}}{I_S \mp \Delta I_S}\right) \\ &= V_T \ln\left(\frac{I_{bias}}{I_S}\right) - V_T \ln\left(1 \mp \frac{\Delta I_S}{I_S}\right) \end{aligned} \tag{B.3}$$

which gives

$$\pm \Delta V_{BE} = -V_T \ln\left(1 \mp \frac{\Delta I_S}{I_S}\right). \tag{B.4}$$

For $\Delta V_{BE} = \pm 1.6 V_T = 41.36m\text{V}$, (B.4) results then

$$a) \quad \exp(-1.6) = 1 - \frac{\Delta I_S}{I_S} \Rightarrow \frac{\Delta I_S}{I_S}|_{min} = 0.8 \tag{B.5}$$

$$b) \quad \exp(+1.6) = 1 + \frac{\Delta I_S}{I_S} \Rightarrow \frac{\Delta I_S}{I_S}|_{max} = 3.95 \tag{B.6}$$

In order to compensate the spread in I_S of $\frac{3.95}{0.8} = 4.95 \approx 5$, a minimum current ratio of 5 is required for I_{bias}, which is the coarse trimming capability proposed. The fine trimming allows an extra margin, since the bitstream controlled current source can be programmed from 0 up to I_{bias}.

Appendix C

Second- and third-order incremental $\Delta\Sigma$ modulator ENOB

C.1 Second-order $\Delta\Sigma$ modulator

C.1.1 Discrete-time model

The two integrators inputs V_A and V_B, as shown inf Fig. C.1(a) or (b), where the equivalent model is presented, are given by

$$V_A[n] = V_{in}[n] - bs[n]V_{REF} \tag{C.1}$$

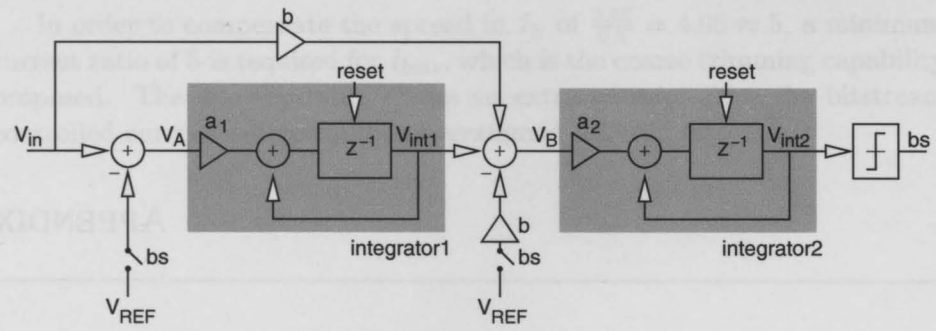
$$V_B[n] = V_{int1} + b\left(V_{in}[n] - bs[n]V_{REF}\right). \tag{C.2}$$

while the two outputs given by

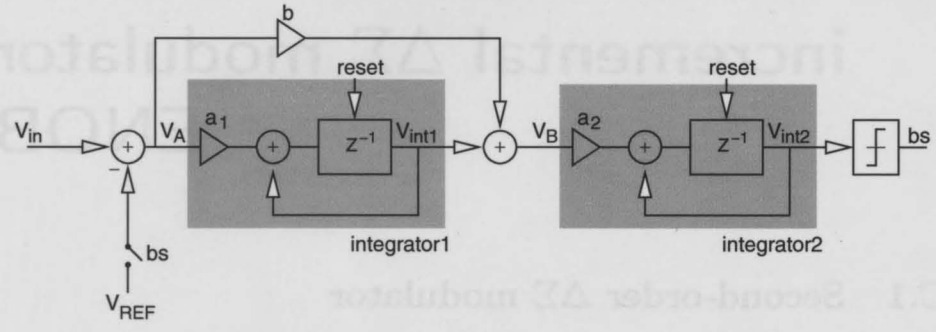
$$\begin{aligned} v_{int1}[n] &= v_{int1}[n-1] + a_1 V_A[n-1] \\ &= v_{int1}[n-1] + a_1\left(V_{in}[n-1] - bs[n-1]V_{REF}\right) \end{aligned} \tag{C.3}$$

and

$$\begin{aligned} v_{int2}[n] &= v_{int2}[n-1] + a_2 V_B[n-1] \\ &= v_{int2}[n-1] + a_2 V_{int1}[n-1] + \\ &\quad a_2 b\left(V_{in}[n-1] - bs[n-1]V_{REF}\right). \end{aligned} \tag{C.4}$$



(a)



(b)

Figure C.1: (a) Discrete-time model of the second-order ΔΣ modulator and (b) its equivalent model.

C.1.2 Resolution - Effective number of bits (ENOB)

The output of the first integrator after n cycles is given by

$$v_{int1}[n] = a_1 \sum_{k=0}^{n-1} (V_{in}[k] - bs[k]V_{REF}) \tag{C.5}$$

where $bs[k] = 0, 1$ is the comparator output, i.e. the bitstream value of former cycles. The output of the second integrator after n cycles after reset ($v_{int2}[0] = 0$) is given by

$$\begin{aligned} v_{int2}[n] &= a_2 \sum_{l=0}^{n-1} v_{int1}[l] + a_2 b \sum_{k=0}^{n-1} (V_{in}[k] - bs[k]V_{REF}) \\ &= a_2 \sum_{l=0}^{n-1} a_1 \sum_{k=0}^{l-1} (V_{in}[k] - bs[k]V_{REF}) + \end{aligned}$$

$$a_2 b \sum_{k=0}^{n-1} (V_{in}[k] - bs[k]V_{REF}) \tag{C.6}$$

Assuming once again $V_{in}[k] = V_{in}$, (C.6) can be rewritten as

$$v_{int2}[n] = a_2 \left(\frac{1}{2}a_1(n-1)n + bn\right) V_{in} - a_2 V_{REF} D_2[n] \tag{C.7}$$

where

$$D_2[n] = a_1 \sum_{l=0}^{n-1} \sum_{k=0}^{l-1} bs[k] + b \sum_{k=0}^{n-1} bs[k]. \tag{C.8}$$

Rearranging equation (C.7) results in

$$\begin{aligned} \frac{v_{int2}[n]}{a_2\left(\frac{1}{2}a_1(n-1)n + bn\right)} &= V_{in} - \frac{1}{\left(\frac{1}{2}a_1(n-1)n + bn\right)} V_{REF}.D_2[n] \\ &= V_{in} - \hat{V}_{in} \end{aligned} \tag{C.9}$$

where

$$\hat{V}_{in} = \frac{1}{\left(\frac{1}{2}a_1(n-1)n + bn\right)} V_{REF} D_2[n] \tag{C.10}$$

is an estimate for V_{in}. Assuming $\pm\frac{V_{REF}}{2}$ as a bound for $V_{int2}[n]$, also assuming the loop stable for all inputs (what does not necessarily stands for all values of a_2, a_1 and b due to stability issues), the limits on the error for the estimate of V_{in} can be written as

$$\begin{aligned} -\tfrac{V_{REF}}{2} \tfrac{1}{a_2\left(\frac{1}{2}a_1(n-1)n+bn\right)} \leq \\ V_{in} - \tfrac{1}{\left(\frac{1}{2}a_1(n-1)n+bn\right)} V_{REF} D_2[n] \\ \leq \tfrac{V_{REF}}{2} \tfrac{1}{a_2\left(\frac{1}{2}a_1(n-1)n+bn\right)}. \end{aligned} \tag{C.11}$$

From these limits, the equivalent LSB voltage is given by

$$\frac{V_{LSB}}{2} = \frac{V_{REF}}{2} \frac{1}{a_2\left(\frac{1}{2}a_1(n-1)n + bn\right)} \tag{C.12}$$

which allows to calculate the effective number of bits n_{bit}, $ENOB_{2nd,ideal}$ that ideally can be a achieved with a second-order A/D converter ($V_{LSB} = \frac{V_{REF}}{2^{nbit}}$):

$$ENOB_{2nd,ideal} = \log_2\left(a_2\left(\frac{1}{2}a_1(n-1)n + bn\right)\right), \tag{C.13}$$

or, if $a_1 = a_2 = b = 1$, as

$$ENOB_{2nd,ideal} \approx 2\log_2(n) - 1 \qquad (n \gg 1). \tag{C.14}$$

C.2 Third-order ΔΣ modulator

C.2.1 Discrete-time model

The three integrators inputs V_A, V_B and V_C of a third-order ΔΣ modulator, as shown in Fig.C.2(a) or in Fig.C.2(b), where its equivalent model is presented, are given by

$$V_A[n] = V_{in}[n] - bs[n].V_{REF}, \tag{C.15}$$

$$V_B[n] = V_{int1}[n] + b\left(V_{in}[n] - bs[n]V_{REF}\right), \tag{C.16}$$

and

$$V_C[n] = V_{int2}[n] + b_2\left(V_{in}[n] - bs[n]V_{REF}\right). \tag{C.17}$$

while the three integrators outputs are given by

$$\begin{aligned} v_{int1}[n] &= v_{int1}[n] + a_1 V_A[n] \\ &= v_{int1}[n] + a_1\left(V_{in}[n] - bs[n]V_{REF}\right), \end{aligned} \tag{C.18}$$

$$\begin{aligned} v_{int2}[n] &= v_{int2}[n-1] + a_2 V_B[n-1] \\ &= v_{int2}[n-1] + a_2 V_{int1}[n-1] + \\ &\quad a_2 b\left(V_{in}[n-1] - bs[n-1]V_{REF}\right) \end{aligned} \tag{C.19}$$

and

$$\begin{aligned} v_{int3}[n] &= v_{int3}[n-1] + a_3 V_C[n-1] \\ &= v_{int3}[n-1] + a_3 V_{int2}[n-1] + \\ &\quad a_3 b_2\left(V_{in}[n-1] - bs[n-1]V_{REF}\right). \end{aligned} \tag{C.20}$$

C.2.2 Effective number of bits (ENOB)

Following the same procedure of the last section, the output of the first integrator after n cycles after reset ($v_{int1}[0] = 0$)) is

$$v_{int1}[n] = a_1 \sum_{k=0}^{n-1} \left(V_{in}[k] - bs[k]V_{REF}\right) \tag{C.21}$$

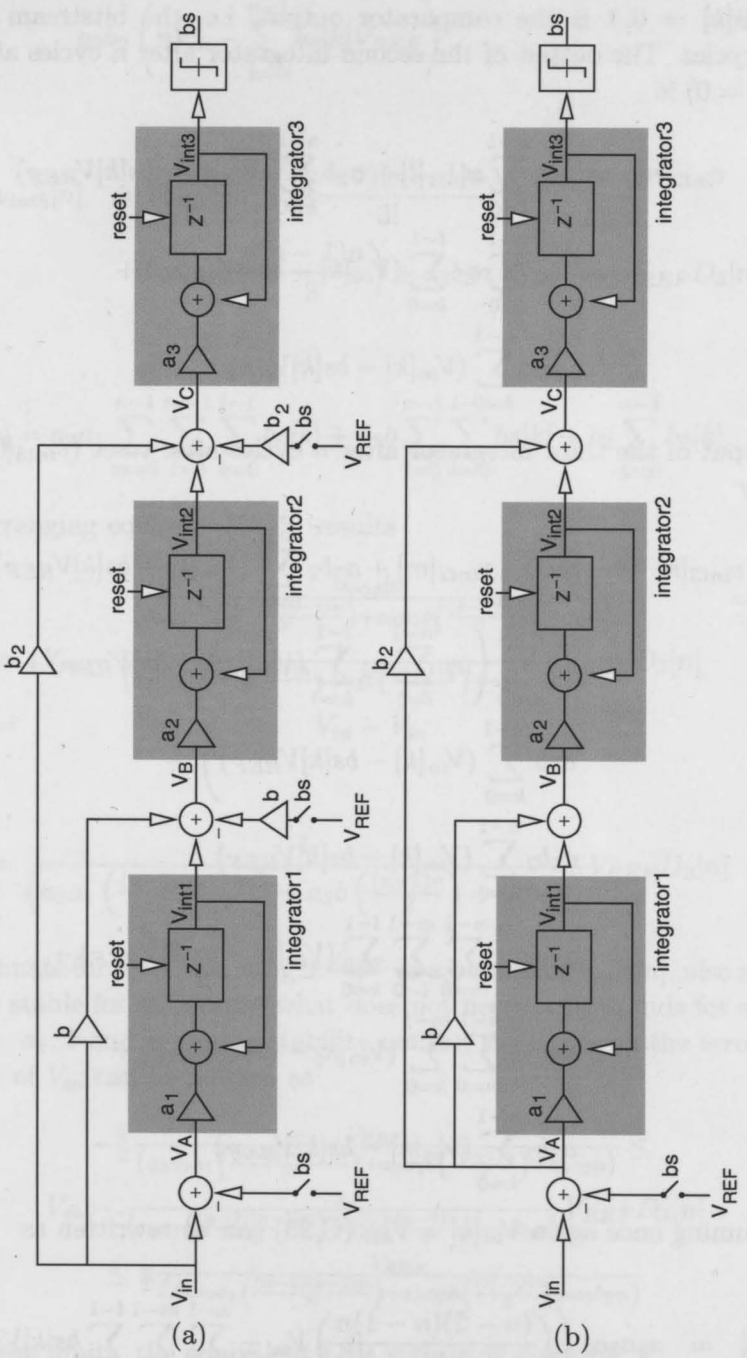


Figure C.2: (a) Discrete-time model of a third-order ΔΣ modulator and (b) its equivalent model.

where $bs[k] = 0, 1$ is the comparator output, i.e, the bitstream value of former cycles. The output of the second integrator after n cycles after reset ($v_{int2}[0] = 0$) is

$$\begin{aligned} v_{int2}[n] &= a_2 \sum_{l=0}^{n-1} v_{int1}[l] + a_2 b \sum_{k=0}^{n-1} (V_{in}[k] - bs[k]V_{REF}) \\ &= a_2 \sum_{l=0}^{n-1} a_1 \sum_{k=0}^{l-1} (V_{in}[k] - bs[k]V_{REF}) + \\ &\quad a_2 b \sum_{k=0}^{n-1} (V_{in}[k] - bs[k]V_{REF}) \end{aligned} \tag{C.22}$$

The output of the third integrator after n cycles after reset ($v_{int3}[0] = 0$) is given by

$$\begin{aligned} v_{int3}[n] &= a_3 \sum_{m=0}^{n-1} v_{int2}[m] + a_3 b_2 \sum_{m=0}^{n-1} (V_{in}[k] - bs[k]V_{REF}) \\ &= a_3 \sum_{m=0}^{n-1} \left(a_2 \sum_{l=0}^{m-1} a_1 \sum_{k=0}^{l-1} (V_{in}[k] - bs[k]V_{REF}) + \right. \\ &\quad \left. a_2 b \sum_{k=0}^{m-1} (V_{in}[k] - bs[k]V_{REF}) \right) + \\ &\quad a_3 b_2 \sum_{k=0}^{n-1} (V_{in}[k] - bs[k]V_{REF}) \\ &= a_3 a_2 a_1 \sum_{m=0}^{n-1} \sum_{l=0}^{m-1} \sum_{k=0}^{l-1} (V_{in}[k] - bs[k]V_{REF}) + \\ &\quad a_3 a_2 b \sum_{m=0}^{n-1} \sum_{k=0}^{m-1} (V_{in}[k] - bs[k]V_{REF}) + \\ &\quad a_3 b_2 \sum_{k=0}^{n-1} (V_{in}[k] - bs[k]V_{REF}) \end{aligned} \tag{C.23}$$

Assuming once again $V_{in}[k] = V_{in}$, (C.23) can be rewritten as

$$\begin{aligned} v_{int3}[n] &= a_3 a_2 a_1 \left(\left(\frac{(n-2)(n-1)n}{3!} \right) V_{in} - \sum_{m=0}^{n-1} \sum_{l=0}^{m-1} \sum_{k=0}^{l-1} bs[k]V_{REF} \right) + \\ &\quad a_3 a_2 b \left(\left(\frac{(n-1)n}{2} \right) V_{in} - \sum_{m=0}^{n-1} \sum_{k=0}^{m-1} bs[k]V_{REF} \right) + \end{aligned}$$

$$a_3 b_2 \left(nV_{in} - \sum_{k=0}^{n-1} bs[k] V_{REF} \right) \tag{C.24}$$

or as

$$\begin{aligned} v_{int3}[n] &= \left(a_3 a_2 a_1 \left(\frac{(n-2)(n-1)n}{3!} \right) + \right. \\ &\quad \left. a_3 a_2 b \left(\frac{(n-1)n}{2} \right) + a_3 b_2 n \right) V_{in} - a_3 V_{REF} D_3[n] \end{aligned} \tag{C.25}$$

where

$$D_3[n] = a_2 a_1 \sum_{m=0}^{n-1} \sum_{l=0}^{m-1} \sum_{k=0}^{l-1} bs[k] + a_2 b \sum_{l=0}^{n-1} \sum_{k=0}^{l-1} bs[k] + b_2 \sum_{k=0}^{n-1} bs[k]. \tag{C.26}$$

Rearranging equation (C.25) results

$$\begin{aligned} & \frac{v_{int3}[n]}{\left(a_3 a_2 a_1 \left(\frac{(n-2)(n-1)n}{3!}\right) + a_3 a_2 b \left(\frac{(n-1)n}{2}\right) + a_3 b_2 n\right)} = \\ &= V_{in} - \frac{1}{\left(a_2 a_1 \left(\frac{(n-2)(n-1)n}{3!}\right) + a_2 b \left(\frac{(n-1)n}{2}\right) + a_3 b_2 n\right)} V_{REF} D_3[n] \\ &= V_{in} - \hat{V}_{in} \end{aligned}$$

where

$$\hat{V}_{in} = \frac{1}{\left(a_2 a_1 \left(\frac{(n-2)(n-1)n}{3!}\right) + a_2 b \left(\frac{(n-1)n}{2}\right) + a_3 b_2 n\right)} V_{REF} D_3[n] \tag{C.27}$$

is an estimate for V_{in}. Assuming $\pm \frac{V_{REF}}{2}$ as a bound for $V_{int3}[n]$, also assuming the loop stable for all inputs (what does not necessarily stands for all values of a_3, a_2, a_1, b and b_1 due to stability issues), the limits on the error for the estimate of V_{in} can be written as

$$\begin{aligned} & -\tfrac{1}{2} \frac{V_{REF}}{\left(a_3 a_2 a_1 \left(\frac{(n-2)(n-1)n}{3!}\right) + a_3 a_2 b \left(\frac{(n-1)n}{2}\right) + a_3 b_2 n\right)} \leq \\ V_{in} - & \frac{1}{\left(a_2 a_1 \left(\frac{(n-2)(n-1)n}{3!}\right) + a_2 b \left(\frac{(n-1)n}{2}\right) + a_3 b_2 n\right)} V_{REF} D_3[n] \\ & \leq \tfrac{1}{2} \frac{V_{REF}}{\left(a_3 a_2 a_1 \left(\frac{(n-2)(n-1)n}{3!}\right) + a_3 a_2 b \left(\frac{(n-1)n}{2}\right) + a_3 b_2 n\right)} \end{aligned}$$

From these limits, the equivalent LSB voltage is given by

$$\frac{V_{LSB}}{2} = \frac{1}{2} \frac{V_{REF}}{2^{nbit}} = \frac{1}{2} \frac{V_{REF}}{\left(a_3 a_2 a_1 \left(\frac{(n-2)(n-1)n}{3!}\right) + a_3 a_2 b \left(\frac{(n-1)n}{2}\right) + a_3 b_2 n\right)}$$

which allows to calculate the effective number of bits n_{bit}, $ENOB$ as

$$ENOB_{3rd,ideal} = \log_2\left(a_3a_2a_1\left(\frac{(n-2)(n-1)n}{3!}\right)+ a_3a_2b\left(\frac{(n-1)n}{2}\right)+a_3b_2n\right) \quad \text{(C.28)}$$

or, if $a_3 = a_2 = a1 = 1$, as

$$ENOB_{3rd,ideal} \approx 3\log_2(n) - 2.58 \qquad (n \gg 1). \quad \text{(C.29)}$$

APPENDIX D

Design of a third-order CIFB-IF $\Delta\Sigma$ modulator

The high-pass transfer function of the quantization noise NTF is actually the starting point of any loop filter synthesis procedure. The $NTF(z)$ is designed to have all zeroes at $z = 1$, e.g.

$$NTF(z) = \frac{(z-1)^3}{D(z)}, \tag{D.1}$$

so in practice the converter can be used for various oversampling ratios, which might be the case depending on the sensor applications.

While the numerator of the $NTF(z)$ set the zeroes, the function $D(z)$ of the NTF, set the poles of the system, which have to ensure stability of the modulator, while shaping as much quantization noise away from DC as possible. Using a digital filter software to design a third-order high-pass digital filter (e.g. synthesizeNFT Matlab [71]) where the pass-band is adjusted until the peak of NTF is less than 1.4 $|NTF(z)| < 1.4$, which guarantees stability [91][72], the following NTF is found

$$NTF(z) = \frac{(z-1)^3}{z^3 - 2.3741z^2 + 1.9294z - 0.75321}. \tag{D.2}$$

Substituting (D.2) in $H(z)$

$$H(z) = \frac{1 - NTF(z)}{NTF(z)} \tag{D.3}$$

gives[1]

[1] Interesting to see that the zeroes of the $NTF(z)$ are the poles of $H(z)$.

$$H(z) = \frac{0.6259z^2 - 1.0706z + 0.4679}{(z-1)^3}. \tag{D.4}$$

The next step is to choose the structure of the modulator, in this case CIFB-IF topology, as previously shown in Fig. 4.14(a). This topology with the input feed-forwarded is the same of that discussed for a second-order, obviously adapted for a structure with one additional integrator. The benefit is the same: it allows the implementation of the charge balancing discussed previously, since it can be modified to an equivalent structure, as shown in Fig. 4.14(b). But it also helps to reduce the amount of the signal $X(z)$ seem at the output of intermediate integrators (first and second ones).

From the architecture, the two loop-filter functions $L_1(z)$ and $L_0(z)$ can be derived:

$$\begin{aligned} L_1(z) &= \frac{1}{z-1}\left[-b_3 + \frac{a_3}{z-1}\left(-b_2 + \frac{a_2}{z-1}(-b_1)\right)\right] \\ &= -\frac{b_3z^2 + (a_3b_2 - 2b_3)z - a_3b_2 + a_3a_2b_1 + b_3}{(z-1)^3} \end{aligned} \tag{D.5}$$

$$L_0(z) = \frac{1}{z-1}\left[c_3 + \frac{a_3}{z-1}\left(c_2 + \frac{a_2}{z-1}a_1\right)\right] \tag{D.6}$$

$$= \frac{c_3z^2 + (a_3c_2 - 2c_3)z - a_3c_2 + a_3a_2c_1 + c_3}{(z-1)^3} \tag{D.7}$$

The $NTF(z) = \frac{1}{1-L1}$ and $STF(z) = \frac{L0}{1-L1}$ are thus given by

$$NTF(z) = \frac{(z-1)^3}{(z-1)^3 + b_3(z-1)^2 + a_3b_2(z-1) + a_3a_2b_1} \tag{D.8}$$

and

$$STF(z) = \frac{c_3(z-1)^2 + a_3c_2(z-1) + a_3a_2a_1}{(z-1)^3 + b_3(z-1)^2 + a_3b_2(z-1) + a_3a_2b_1} \tag{D.9}$$

which requires $a_1 = b_1$ for an $STF(z=1) = 1$.

The coefficients b_i are found by deriving from the architecture, the transfer function from the output $Y(z)$ to $V_{int3}(U(z))$, exactly the function $L1$ and equating it to $-H(z)$ as given in (D.3). By comparison, the following coefficients are found: $b_3 = 0.6259$, $b_2 = 0.1812$ and $b_1 = a_1 = 0.0232$ (with $a2 = a3$ assumed equal to 1). System level simulations are performed to verify the modulator operation, paying attention to the output of the integrators, so the dynamic scaling of coefficients can be performed, which

will ensure that all nodes have approximately the same power level, with no clipping. Using a sine wave input, with amplitude varying from 0.25% to 0.75% of V_{REF}, the coefficients found after scaling and rounding are $b_2 = 0.5$, $b_3 = 0.5$, $b_1 = a_1 = 0.5$ and $a_2 = 0.125$ and $a3 = 0.25$.

APPENDIX E

Weak inversion operation

The drain current of a MOS transistor operating in weak inversion (saturated in weak inversion) is given by [92]

$$I_D = I_{D0} \exp\left(\frac{V_G}{nV_T}\right)\left\{\exp\left(\frac{-V_S}{V_T}\right) - \exp\left(\frac{-V_D}{V_T}\right)\right\} \tag{E.1}$$

where

$$I_{D0} \equiv I_S \exp\left(\frac{-V_{T0}}{n.V_T}\right) \tag{E.2}$$

is the leakage current appearing for $V_G = 0$, written as a function of the so specific current I_S defined as [93]

$$I_S = 2n\mu C_{ox}\left(\frac{W}{L}\right)V_T^2, \tag{E.3}$$

which, at a given temperature, depends essentially on the aspect ratio $\left(\frac{W}{L}\right)$of the device, slope factor n and mobility μ. The parameter I_{D0} is not well controlled since it depends exponentially on V_{T0}. Any device biasing should avoid dependence on such parameter. The voltages V_G, V_S and V_D are the gate, source and drain voltages, respectively, all referred to the bulk voltage (p-substrate or n-well).

For $V_D - V_S > 4V_T$ to $5V_T$, the device is said in forward saturation, and (E.1) simplifies to

$$I_D = I_{D0} \exp\left(\frac{V_G}{nV_T}\right)\exp\left(\frac{-V_S}{V_T}\right) \tag{E.4}$$

since the (reverse) component of the current associated to $V_G - V_D$ is negligible. It is usual to write (E.4) as

$$I_D = I_0\left(\frac{W}{L}\right)\exp\left(\frac{V_G}{nV_T}\right)\exp\left(\frac{-V_S}{V_T}\right) \tag{E.5}$$

where

$$I_0 = 2n\mu C_{ox} V_T^2 \exp\left(\frac{-V_{T0}}{nV_T}\right), \tag{E.6}$$

so the aspect ratio of the device appears explicitly in the expression of the drain current I_D.

The transconductance g_m is given by

$$g_m \equiv \frac{\partial I_D}{\partial V_G} = \frac{1}{nV_T} I_D \tag{E.7}$$

which results in

$$\frac{g_m}{I_D} = \frac{1}{nV_T}. \tag{E.8}$$

APPENDIX F

Integrator's opamp load capacitances

Settling errors are introduced either on sampling or integrating phases. Normally, the capacitive load of the amplifier during the integration phase is larger than the load during the sampling phase [67], so the former phase places the worst-case scenario, and determines the settling time properties of the integrator.

The open-loop capacitance is expressed as

$$C_{OL} = C_L + \frac{C_I(C_S + C_P)}{C_I + (C_S + C_P)} \tag{F.1}$$

where C_S, C_I, C_P and C_L are the sampling, integrating, amplifier's parallel input and output capacitances, respectively. C_L represents the drain capacitance of the transistors in the output stage and all other capacitances physically connected at the integrator's output during the integration phase. C_L has great importance since it adds to the open-loop capacitive load at the output of the integrator.

The closed loop capacitance is given by

$$C_{CL} = \frac{C_{OL}}{G} \tag{F.2}$$

where G is the feed-back factor, which depends on the amplifier's configuration, i.e. if non-inverting

$$G_{non-inverting} = \frac{C_I}{C_I + C_S + C_P} \tag{F.3}$$

or inverting

$$G_{inverting} = \frac{C_I}{C_S + C_P} \tag{F.4}$$

configuration. For example, the closed-loop capacitance of a non-inverting amplifier is given by

$$C_{CL,non-inverting} = C_S + C_P + C_L + \frac{C_L(C_S + C_P)}{C_I}. \quad \text{(F.5)}$$

APPENDIX G

Self-heating

Self-heating in temperature sensors holds for a rise in the sensor's operating temperature due to an internal heating source. Since it modifies the sensor's temperature, it should be prevented or limited to avoid measurement errors. For instance, if the sensor circuitry dissipates a given amount of power, an if this power increases the sensor's temperature, the measurement is likely to be not accurate.

Because the exact rise in the sensor's temperature depends not only on how much power the sensor consumes, but also on other factors like packaging type and the way the package is mounted, the most effective way of handling self-heating is through prevention, by design, keeping the sensor's power consumption limited to a given level, so the rise in temperature is not observable or, if observable, negligible for a particularly application accuracy.

An attempt to estimate worst-case scenarios due to self-heating and associated measurement errors can be made from the die-to-ambient thermal resistance θ_{JA}(K/W). This means that, to prevent a circuit temperature rise ΔT, a maximum power $P = \frac{\Delta T}{\theta_{JA}}$ can be consumed [22]. Typical values of θ_{JA} for the most common plastic packages are 200K/W [94][95] meaning that a power consumption not greater than 50μW is required to prevent a temperature rise of 0.01°C, for example. Better scenarios are expected if the allowed ΔT is higher, or if the sensor is mounted in contact with a thermal sink. For instance, in this situation, for the same package, θ_{JA} is 50K/W, and a power consumption as great as 200μW is allowed for the same temperature rise. In this case, however, an extra thermal delay was added to the system due to the thermal sink. Anyway, self-heating should be prevented, otherwise the measurement's accuracy might end depending on external factors.

In this section the self-heating is investigated. Not because self-heating is expected to be present in the actual sensor, but because in one par-

Table G.1: Typical values for the thermal resistance θ_{JA}(junction to ambient) of some packages for two different mounting condition.

θ_{JA}[°C/W]	still air	clamped to a heat sink
TO46 metal	400	24
TO-92 plastic	160	55
Ceramic DIL 8-28pins	110-55	-
SO-8	220	55

Table G.2: Supply voltage, measured current and estimated power per device, and measured temperature rise (in °C) due to self-heating.

V_{DD}(V)	I_{CC}(mA)	P_{diss}(mW)	ΔT(°C) @ 30°C
2.5	0.06	≈ 0.15	-
3.3	0.25	≈ 1	-
5.5	0.92	≈ 5	0.1

ticular taped-out design, a chain of two inverters with floating input was added. This chain does not present any function, but because of the current $I_{CC} = V_{DD}/(R_{on,p} + R_{on,n})$ that flows through them, from V_{DD} to ground, it acts as a heating source inside the sensor, and the self-heating effect can be estimated.

The procedure to evaluate the self-heating is as follows. Because of the inverter area, appropriate values for I_{CC} (and P_{diss}) were established for $V_{DD} = 2.5 - 5.5$V (and for the PMOS and NMOS transistors sizes used in the inverters). These values are summarized in G.2. The chain of inverters was placed very close to the PNP transistors and PNP biasing CMOS current sources (the most temperature sensitive circuits).

The analysis was as follows. Four sensors (with the chain of inverters inside each) were measured over temperature for $V_{DD} = 2.5 - 5.5$V. These measurements[1] (for devices 0 and 2) are shown in Figure G.1(a). From these curves it is already clear that self-heating is present when $V_{DD} = 5.5$V, which appears as an up-shift on the error when compared with the error for $V_{DD} = 2.5 - 3.3$V. Although part of the effect could be also attributed to a limited PSRR, it is excessively high for the actual design.

After the initial characterization, the heating sources of devices 0 and 1 were removed (through laser blasting), and measurements were performed again. The results (for devices 0 and 2) are shown in Figure G.1(b).

For $V_{DD} = 2.5$V and 3.3V, the temperature error for all devices is mainly

[1] batch calibrated.

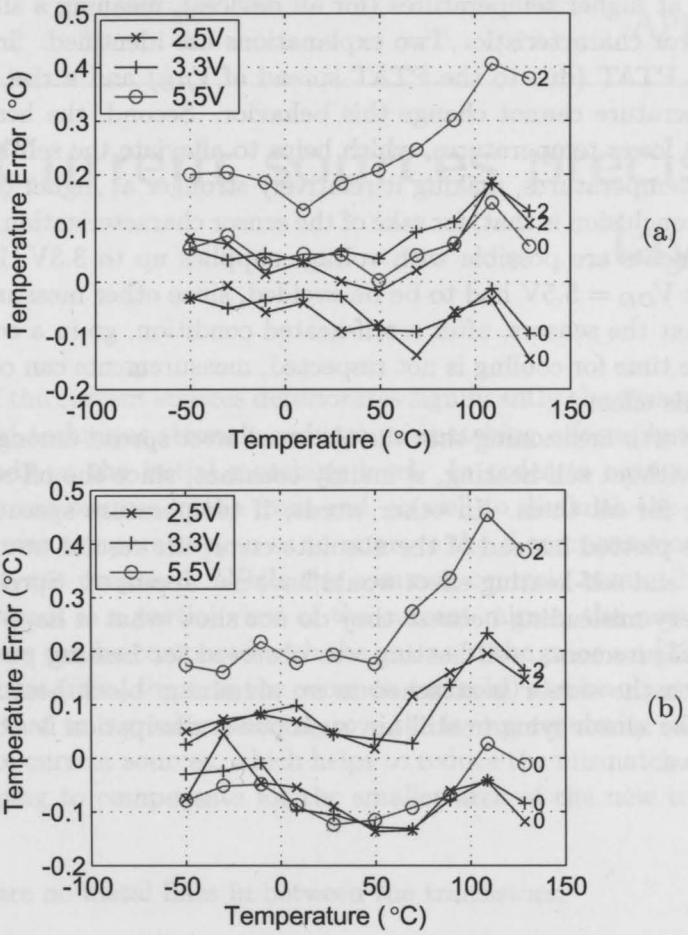


Figure G.1: Temperature error for devices 0 and 2. (a) Situation where both circuits have the heating source inside. (b) Situation where device 0 does not have the heating source anymore, while device 2 remains unchanged.

the same[2] values initially observed. However, for $V_{DD} = 5.5$V, the strong up-shift is not more visible for circuits 0 and 1 (whose heating sources were removed). This clearly confirms the effect of self-heating.

Self-heating in this sense appears primarily as an off-set in the measurements, meaning a rise in the internal temperature, as expected. This off-set is higher at higher temperatures (for all devices), meaning a slight rotation in the error characteristic. Two explanations are identified: first, the error spread is PTAT (due to the PTAT spread of V_{BE}) and a rise in the internal temperature cannot change this behavior. Second, the heat-transfer is helped at lower temperatures, which helps to alleviate the self-heating effect at lower temperatures, making it relatively stronger at higher ones.

The conclusion is that, for sake of the sensor characterization, appropriate measurements are possible with voltage supplies up to 3.3V. Important to note that $V_{DD} = 5.5$V had to be be avoided, since other measurements have shown that the sensors, after a self-heated condition, go in a cooling phase, and if the time for cooling is not respected, measurements can contain errors due to this effect.

It is worth mentioning that the observed error spread among the devices, with or without self-heating, is mainly constant, since the off-set is roughly the same for all them. In other words, if temperature spread among the devices is plotted instead of the absolute error, the results would have been the same and self-heating effect would have not appeared. Spread curves are simply very misleading because they do not show what is happening.

In measurements, self-heating was observed for heating powers close to mW, when the sensor is attached to an aluminum block heating sink. Assuming the sensor lying in still air, such power dissipation is expected to be prohibitive.

[2]For $V_{DD} = 3.3$V, a weak self-heating effect (below 0.05°C) is visible. The author would like to interpret 3.3V as the limit supply voltage for self-heating, above which self-heating is not more negligible.

APPENDIX H

Current sources precision layout

Mismatch of the current sources deteriorates significantly the sensor's performance. DEM technique strongly reduces mismatching effects, but the final result depends on the initial matching level. In order to reduce the mismatch of current sources in the front-end, as well to diminish the mirroring dependence over temperature, so as to improve the β-compensation scheme, a new layout was proposed. While the common centroid cannot be seen in Fig. H.1, since it is a partial view of the current mirror, the new layout of current sources can be compared with the one as in the prior art [22][1]. Three main differences (in addition to the common centroid) can be observed in the new layout, that to a great extent improve the layout regularity (and layer's continuity) of current sources, which helps to reduce the mismatch of current sources, helping to compensate for the smaller area of the new transistors. They are:

- there are no metal lines in between the transistors;
- the gates of the both transistors are continuous and;
- the layout of the current sources are not mirrored.

The absence of metal lines in between the transistors keeps the region among the transistors more uniform, which helps to reduce mismatch. This is in addition to the fact that without the metal lines the transistors are closer and controlled by continuous gate lines. Since the new transistors are thinner and there is no metal line in between, the gate resistance is not of concern

[1]The layout was kindly provided by the author.

Table H.1: Comparison between sizes of the main transistor used in the current sources. T1 is the main current source transistor while T2 is the cascode one.

	T1 (mirror transistor)		T2 (cascode transistor)	
	as in [22]	this work	as in [22]	this work
$W(\mu m)$	10	4.4	10	4.4
$L(\mu m)$	40	64	1	1
$A(\mu m^2)$	400	≈ 282	10	4.4

(considering the gates are short-circuited by external metal lines). Finally, because the layout mirroring is not used now (which in general has to be avoided [35]), better matching level is expected in the latter case.

Table H.1 summarizes the sizes of the transistors used. Despite of a smaller area, the new transistors seem to perform better. This can be explained, as just mentioned, by the more uniform layout of the active area, with continuous poly-silicon lines and without metal lines in between the transistors.

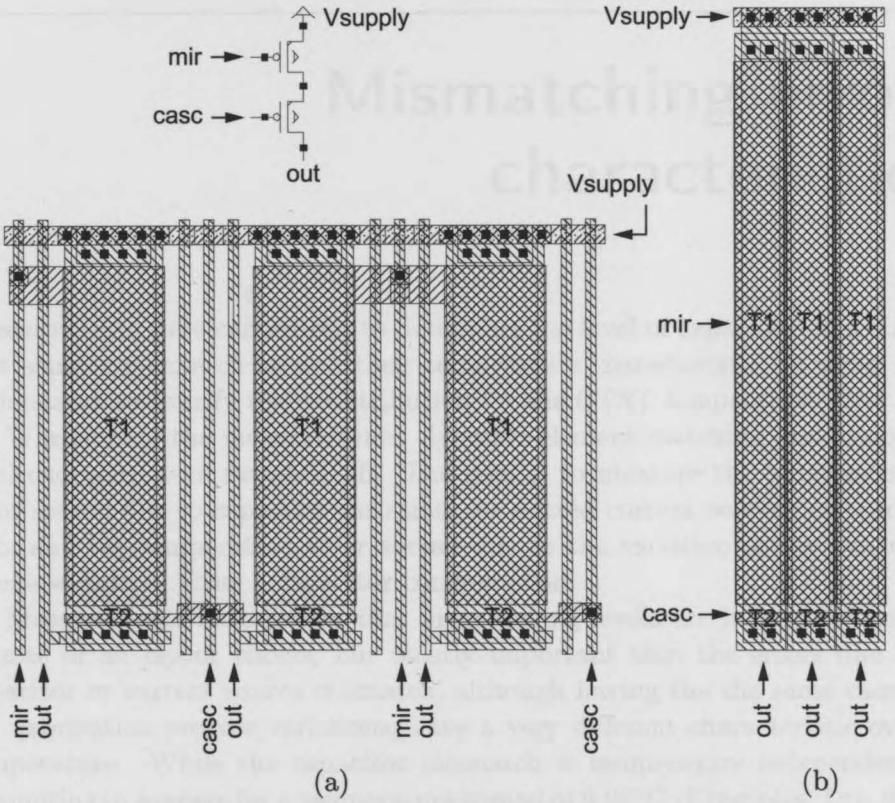


Figure H.1: Partial view of the current mirror in the sensor front-end where three PMOS cascode current sources are shown. (a) Layout as in the prior art [22] and (b) new layout as in the last design presented in this thesis.

APPENDIX I

Mismatching error characteristic

Measurements were carried out to determine the level of capacitors and current sources mismatch as well their temperature characteristic, in order to understand and verify their contribution for the CTAT temperature error.

When doing the measurements, dynamic element matching and system level chopping were switched off. The goal is to measure the temperature error spread due to capacitor variation, for a fixed current source configuration, and the temperature error spread due to the variation of the current source variation, given a capacitor configuration.

Measurements have shown that mismatching levels are indeed different despite of all layout efforts, but mostly important that the errors due to capacitor or current source mismatch, although having the the same cause, i.e. fabrication process variations, have a very different characteristic over temperature. While the capacitor mismatch is temperature independent, accounting in average for a temperature spread of 0.06°C (5 samples, 1σ), the current source mismatch is temperature dependent (strongly CTAT, which agrees with the theory developed in Section 3.8 of Chapter 3) and worse than the capacitor mismatch, accounting in average for a temperature spread of 0.6°C, 0.45°C and 0.25°C (1σ) at −55°C, 30°C and 125°C, respectively. Figure I.1 shows the error (and measurement spread over temperature) in the temperature reading when the capacitors are changed, for a fixed current source configuration (system level chopping is off) while Fig. I.2 shows the error (and spread) in the temperature reading when the current sources are changed, given a fixed capacitor configuration (system level chopping is off).

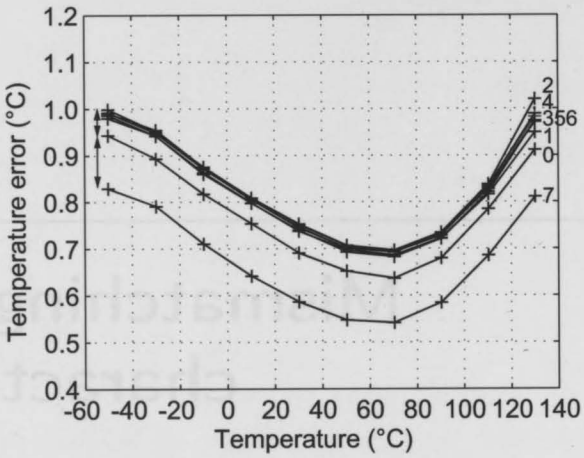


Figure I.1: Measured temperature error for one sample selected at random due to capacitor mismatch ($\alpha = 8$) for a fixed current source configuration without DEM (system level chopping is off).

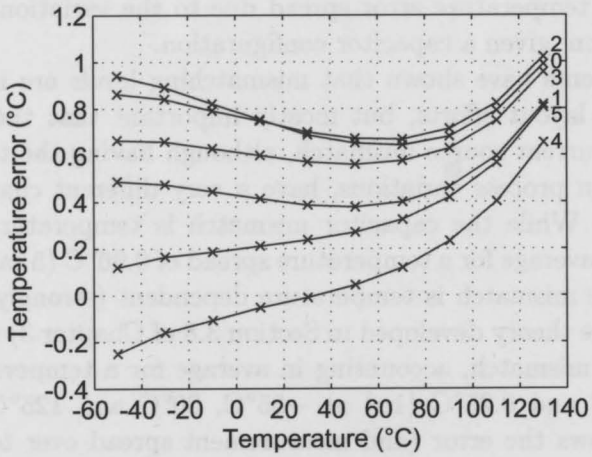


Figure I.2: Measured temperature error for one random sensor due to current source mismatch ($1 + m$ current sources, $m = 5$) for a fixed capacitor configuration (system level chopping is off).

Bibliography

[1] J. H. Huijsing, "Smart sensor systems. why? where? how?" in *Smart Sensor Systems*, G. C. M. Meijer, Ed. John Wiley and Sons, 2008, ch. 1, pp. 1–21.

[2] G. C. M. Meijer, "Smart temperature sensors and temperature-sensor systems," in *Smart Sensor Systems*, G. C. M. Meijer, Ed. John Wiley and Sons, 2008, ch. 7, pp. 185–223.

[3] R. A. Bianchi *et al.*, "CMOS-compatible smart temperature sensors," *Microelectronis Journal, Elsevier*, vol. 29, pp. 627–636, 1998.

[4] ——, "CMOS-compatible temperature sensor with digital output for wide temperature range applications," *Microelectronis Journal, Elsevier*, vol. 31, pp. 803–810, 2000.

[5] A. P. Chandrakasan *et al.*, "Design considerations for ultra-low energy wireless microsensor nodes," *IEEE Transactions on Computers*, vol. 54, no. 6, pp. 727–740, 2005.

[6] C. Kolle *et al.*, "Ultra low-power monolithically integrated capacitive pressure sensor for tire pressure monitoring," in *Proc. IEEE Sensors*, Oct. 2004, pp. 244–247.

[7] G. C. M. Meijer, "Concepts and focus point for intelligent sensor systems," *Sensors and Actuators*, vol. 41, pp. 183–191, 1994.

[8] M. A. P. Pertijs, K. A. A. Makinwa, and J. H. Huijsing, "A CMOS smart temperature sensor with a 3σ inaccuracy of 0.1°C from −55°C to 125°C," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 12, pp. 2805–2815, Dec. 2005.

[9] A. L. Aita, M. A. P. Pertijs, K. A. A. Makinwa, and J. H. Huijsing, "A CMOS smart temperature sensor with a batch-calibrated inaccuracy of ± 0.25°C (3σ) from −70°C to 130°C," in *Proc. IEEE International*

Solid-State Circuits Conference - Digest of Technical Papers, Feb. 2009, pp. 342–343(343a).

[10] K. A. A. Makinwa, "Temperature sensor performance survey 1996-2010," 2010. [Online]. Available: http://www.tudelft.nl

[11] P. R. Kinget, "Device mismatch and tradeoffs in the design of analog circuits," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 6, pp. 1212–1224, 2005.

[12] F. Sebastiano *et al.*, "A 1.2V 10μW NPN-based temperature sensor in 65nm CMOS with an inaccuracy of ± 0.2°C (3σ) from -70°C to 125°C," in *Proc. IEEE International Solid-State Circuits Conference - Digest of Technical Papers*, Feb. 2010, pp. 9–10.

[13] K. Souri, M. Kashmiri, and K. Makinwa, "A CMOS temperature sensor with an energy-efficient zoom ADC and an inaccuracy of ± 0.25°C (3σ) from -40°C to 125°C," in *Proc. IEEE International Solid-State Circuits Conference - Digest of Technical Papers*, Feb. 2010, pp. 6–7.

[14] B. Boser and B. A. Wooley, "The design of sigma-delta modulation analog-do-digital converters," *IEEE Journal of Solid-State Circuits*, vol. 23, no. 6, pp. 1298–1308, Dec. 1988.

[15] G. Wang, "CMOS bandgap references and temperature sensors and their applications," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, Jan. 2005.

[16] W. Bracke, "Generic architectures and design methods for autonomous sensors," Ph.D. dissertation, Katholieke Universiteit Wetenschappen, Leuven, Belgium, Nov. 2006.

[17] "International temperature scale of 1990 - ITS-90." [Online]. Available: http://www.its-90.com/

[18] C. P. L. van Vroonhoven and K. A. A. Makinwa, "A CMOS temperature-to-digital converter with an inaccuracy of ± 0.5°C (3σ) from -55°C to 125°C," in *Proc. IEEE International Solid-State Circuits Conference - Digest of Technical Papers*, Feb. 2008, pp. 576–577, 637.

[19] ——, "Thermal diffusivity sensors for wide-range temperature sensing," in *Proc. IEEE Sensors*, Oct. 2008, pp. 764–767.

[20] I. M. Filanovsky and S. T. Lim, "Temperature sensor application of diode-connected mos transistors," in *Proc. IEEE International Symposium on Circuits and Systems*, vol. V.2, May 2002, pp. 149–152.

[21] I. M. Filanovsky and A. Allam, "Mutual compensation of mobility and threshold voltage temperature effects with applications in CMOS circuits," *IEEE Transactions on Circuits and Systems - I: Fundamental Theory and Applications*, vol. 48, no. 7, pp. 876–884, Jul. 2001.

[22] M. A. P. Pertijs, "High accuracy temperature sensors," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, Nov. 2005.

[23] 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.

[24] A. Bakker, "High-accuracy CMOS smart temperature sensors," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, Apr. 2000.

[25] K. R. Laker and W. M. C. Sansen, *Design of Analog Integrated Circuits and Systems.* Singapore: McGraw-Hill, 1994.

[26] T. C. Verster, "P-n junction as an ultralinear calculable thermometer," *Electronics Letters*, vol. 4, no. 9, pp. 175–176, May 1968.

[27] 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.

[28] "AMI Semiconductor documentation," Electrical Parameters CMOS 0.7μm CDMA - under NDA, Belgium, Nov. 2002.

[29] "AMS Austria Micro Systems documentation," Process and Electrical Parameters CMOS 0.35μm C35 - under NDA, Austria, Dec. 2008.

[30] P. R. Gray and R. G. Meyer, *Analisys and Design of analog Integrated Circuits, Third Edition.* New York: John Wiley and Sons, 1993.

[31] G. Massobrio and P. Antognetti, *Semiconductor Device Modeling with SPICE, Second Edition.* New York: McGraw-Hill Inc, 1993.

[32] G. C. M. Meijer, "Concepts for bandgap references and voltage measurement systems," in *Advanced Analog Circuit Design*, J. H. Huijsing *et al.*, Eds. Boston: Kluwer Academic Publishers, 1996, pp. 243–268.

[33] M. A. P. Pertijs, A. L. Aita, K. A. A. Makinwa, and J. H. Huijsing, "Voltage calibration of smart temperature sensors," in *Proc. IEEE Sensors*, Oct. 2008, pp. 756–759.

[34] F. Fruett, "The piezojunction effect in silicon, its consequences and applications for integrated circuits and sensors," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, Sep. 2001.

[35] A. Hastings, *The Art of Analog Layout.* New Jersey: Prentice Hall, 2001.

[36] K. B. Klaassen, "Digitally controlled absolute voltage division," *IEEE Transactions on Instrumentation and Measurement*, vol. IM-24, no. 2, pp. 106–112, 1975.

[37] 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.

[38] G. C. M. Meijer, Patent Appl. 1 000 222, Apr., 1995.

[39] M. A. P. Pertijs and J. H. Huijsing, "A sigma-delta modulator with bitstream-controlled dynamic element matching," in *Proc. ESSCIRC*, Sep. 2004, pp. 187–190.

[40] M. Tuthill, "A switched-current, switched-capacitor temperature sensor in 0.6μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 7, pp. 1117–1122, Jul. 1998.

[41] B. Razavi, *Design of Analog CMOS Integrated Circuits.* New York, NY: McGraw-Hill, 2001.

[42] P. Malcovati, F. Maloberti, C. Fiocchi, and M. Pruzzi, "Curvature compensated BiCMOS bandgap reference with 1-V supply voltage," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 7, pp. 1076–1081, Jul. 2001.

[43] 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.

[44] A. Bakker and J. H. Huijsing, "Micropower CMOS temperature sensor with digital ouptut," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 7, pp. 937–933, Jul. 1996.

[45] 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, Nov. 1996.

[46] 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.

[47] J. H. Huijsing, *Operational Amplifiers - Theory and Design.* Boston: Kluwer Academic Publishers, 2001.

[48] A. Bakker and J. H. Huijsing, "A CMOS chopper opamp with integrated low-pass filter," in *Proc. ESSCIRC*, Sep. 1997, pp. 200–203.

[49] R. Burt and J. Zhang, "A micropower chopper-stabilized operational amplifier using a SC notch filter with synchronous integration inside the continuous-time signal path," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 12, pp. 2729–2736, Dec. 2006.

[50] R. Schreier, J. Silva, J. Steensgaard, and G. C. Temes, "Design-oriented estimation of thermal noise in switched-capacitor circuits," *IEEE Transactions on Circuits and Systems-I: Regular Papers*, vol. 52, no. 11, pp. 2358–2368, Nov. 2005.

[51] 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. 1433–1439, Oct. 1989.

[52] P. R. Kinget, "Device mismatch: An analog design perspective," in *Proc. IEEE International Symposium on Circuits and Systems*, May 2007, pp. 1245–1248.

[53] 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. International Symposium on Circuits and Systems*, May 2001, pp. 368–371.

[54] M. A. P. Pertijs and J. H. Huijsing, "Bitstream trimming of a smart temperature sensor," in *Proc. IEEE Sensors*, Oct. 2004, pp. 904–907.

[55] G. C. M. Meijer *et al.*, "A three-terminal integrated temperature transducer with microcomputer interfacing," *Sensors and Actuators*, vol. 18, pp. 195–206, Jun. 1989.

[56] A. Marques, "High speed CMOS data converters," Ph.D. dissertation, Katholieke Universiteit Leuven, Leuven, Belgium, Jan. 1999.

[57] P. G. A. Jespers, *Integrated Converters.* New York: Oxford University Press, 2001.

[58] "TMP275 data sheet," Burr-Brown Products from Texas Instruments, Texas, USA, Jun. 2006. [Online]. Available: http://www.ti.com

[59] "AD7641 data sheet," Analog Devices Inc., Oct. 2010. [Online]. Available: http://www.analog.com

[60] D. A. Johns and K. Martin, *Analog Integrated Circuit Design.* New York: John Wiley and Sons, 1997.

[61] S. Rabii and B. A. Wooley, *The Design of Low-Voltage, Low-Power Sigma-Delta Modulators.* Boston: Kluwer Academic Publishers, 1998.

[62] J. D. Maeyer, P. Rombouts, and L. Weyten, "A double-sampling extended-counting ADC," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 3, pp. 411–418, Mar. 2004.

[63] J. Markus, J. Silva, and G. C. Temes, "Theory and applications of incremental $\Sigma\Delta$ converters," *IEEE Transactions on Circuits and Systems - I: Regular Papers*, vol. 51, no. 4, pp. 678–690, Apr. 2004.

[64] J. Markus, P. Deval, V. Quiquempoix, J. Silva, and G. C. Temes, "Incremental delta-sigma structures for DC measurement: an overview," in *Proc. IEEE Custom Integrated Circuits Conference*, Sep. 2006, pp. 41–48.

[65] J. C. Candy and G. C. Temes, Eds., *Oversampling Delta-Sigma Data Converters Theory, Design and Simulation.* New York: IEEE Press, 1992.

[66] F. Maloberti, *Data Converters.* Dordrecht, NL: Springer, 2008.

[67] O. Bajdechi, "Systematic design of sigma-delta analog-to-digital converters," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, Sep. 2003.

[68] J. C. Candy, "An overview of basic concepts," in *Delta-Sigma Data Converters: Theory, Design, and Simulation*, S. R. Norsworthy, R. Schreier, and G. C. Temes, Eds. Wiley-IEEE Press, 1996, ch. 1, pp. 1–43.

[69] V. Quiquempoix, P. Deval, A. Barreto, G. Bellini, J. Markus, J. Silva, and G. C. Temes, "A low-power 22-bit incremental ADC," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 7, pp. 1562–1571, Jul. 2006.

[70] 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.

[71] R. Schreier, "The delta-sigma toolbox version 7.1 delsig.zip software toolbox and user's manual," Dec. 2004. [Online]. Available: http://www.mathworks.com

[72] R. Schreier and G. C. Temes, *Understanding Delta-Sigma Data Converters.* New Jersey: IEEE Press and Wiley-Interscience, 2005.

[73] A. Bakker and J. H. Huijsing, *High-Accuracy CMOS Smart Temperature Sensors.* Boston: Kluwer Academic Publishers, 2000.

[74] A. Gharbiya, T. C. Caldwell, and D. A. Johns, "High-speed oversampling analog-to-digital converters," *Int. J. of High Speed Electronics and Systems*, vol. 15, no. 2, pp. 297–317, Jun. 2005.

[75] J. Silva, U.-K. Moon, J. Steensgaard, and G. C. Temes, "A wideband low-distortion delta-sigma ADC topology," *Elec. Lect.*, vol. 37, no. 12, pp. 737–738, Jun. 2001.

[76] P. E. Allen and D. R. Holberg, *CMOS Analog Circuit Design, Second Edition.* New York: Oxford University Press, 2002.

[77] J. C. Candy, "Decimation for sigma delta modulation," *IEEE Transactions on Communications*, vol. COM-34, no. 1, pp. 72–76, Jan. 1986.

[78] J. F. Witte, "Dynamic offset compensated CMOS amplifiers," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, Dec. 2008.

[79] 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.

[80] 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, no. 6, pp. 1379–1384, Dec. 1990.

[81] E. Sackinger and W. Guggenbuhl, "A high-swing, high-impedance MOS cascode circuit," *IEEE Journal of Solid-State Circuits*, vol. 25, pp. 289–298, Feb. 1990.

[82] O. Choksi and L. R. Carley, "Analysis of switched-capacitor common-mode feedback circuit," *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 50, no. 12, pp. 906–917, 2003.

[83] H. C. Nauta and E. H. Nordholt, "New class of high-performance PTAT current sources," *Electronics Letters*, vol. 21, no. 9, pp. 384–386, 1985.

[84] A. L. Aita and K. A. A. Makinwa, "Low-power operation of a precision CMOS temperature sensor based on substrate PNPs," in *Proc. IEEE Sensors*, Oct. 2007, pp. 856–859.

[85] "ADT7301 data sheet," Analog Devices Inc., Oct. 2005. [Online]. Available: http://www.analog.com

[86] C. P. L. van Vroonhoven and K. A. A. Makinwa, "A thermal-diffusivity-based temperature sensor with an untrimmed inaccuracy of ± 0.2°C (3σ) from -55°C to 125°C," in *Proc. IEEE International Solid-State Circuits Conference - Digest of Technical Papers*, Feb. 2010, pp. 12–13.

[87] K. Souri and K. Makinwa, "A 0.12mm^2 7.7μW micropower temperature sensor with an inaccuracy of ± 0.2°C (3σ) from -30°C to 125°C," in *Proc. ESSCIRC*, Oct. 2010.

[88] M. D. Scott, B. E. Boser, and K. S. J. Pister, "An ultra-low energy ADC for smart dust," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 7, pp. 1123–1129, Jul. 2003.

[89] B. Razavi, *RF Microelectronics*. Upper Saddle River, NJ: Prentice-Hall, 1998.

[90] P. Rombouts and L. Weyten, "A versatile nyquist-rate A/D converter with 16-18 bit performance for sensor readout applications," *Elsevier INTEGRATION, the VLSI Journal*, vol. 39, pp. 48–61, 2005.

[91] K. C. H. Chao, S. Nadeem, W. L. Lee, and C. G. Sodini, "A higher order topology for interpolative modulators for oversampling A/D converters," *IEEE Transactions on Circuits and Systems*, vol. 37, no. 3, pp. 309–318, Mar. 1990.

[92] Y. Tsividis and K. Suyama, "MOSFET modeling for analog circuit CAD: problems and prospects," in *Proc. IEEE Custom Integrated Circuits Conference*, May 1993, pp. 14.1.1–14.1.6.

[93] C. C. Enz, F. Krummenacher, and E. A. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low- current applications," *Analog Integr. Circuits Signal Process.*, vol. 8, no. 1, pp. 83–114, Jul. 1995.

[94] "LM35 data sheet," National Semiconductor Corporation, Nov. 2000. [Online]. Available: http://www.national.com

[95] "Thermal resistance table," Linear Technology. [Online]. Available: http://www.linear.com/designtools/packaging/therresist.pdf

Summary

Low-Power High-Accuracy
Smart Temperature Sensors in CMOS Technology
by André Luiz Aita

This thesis describes the main theory, design and implementation of high-accuracy low-power CMOS smart temperature sensors.

Many temperature sensing elements can be used to build smart temperature sensors. If the sensor is integrated though, the options are more restricted and the bipolar transistor, in particular the CMOS-compatible substrate PNP has been extensively used with success. This success, however, depends on the deep understanding of its temperature behavior and nonidealities for further correction.

This thesis is organized as follows. After the Introduction, Chapter 2 - Temperature Sensing Fundamentals - presents the design background on temperature sensors. The substrate PNP transistor characteristics and behavior are discussed in detail since this transistor is the fundamental building block of the sensor.

Chapter 3 - Front-end presents the design of the analog front-end sensor, showing how to generate accurately important voltages like the PTAT voltage ΔV_{BE} and the CTAT voltage V_{BE}. PNP transistors bias currents and current ratios are deeply analyzed in face of accuracy requirements and power consumption constraints.

In Chapter 4 - $\Delta\Sigma$ A/D Converters, the A/D converter of the smart temperature sensor, i.e. the sensor interface, is analyzed and designed. Such block processes the front-end analog output, i.e. ΔV_{BE} and V_{BE} signals, generating a digital signal, whose value is a digital representation of the temperature. Important accuracy-power consumption tradeoffs are addressed focusing on a low-power though accurate to $\pm 0.1°C$ A/D converter.

In Chapter 5 - Realizations, measurement results of three variants of a temperature sensor are presented and analyzed. The last sensor, which is not coincidentally the sensor with best performance, having an inaccuracy below

$\pm 0.25°C$ ($\pm 3\sigma$) over an extended military temperature range, i.e. 200°C from $-70°C$ to 130°C supplied by a total current of $25\mu A$, at $V_{DD} = 2.5V$ and 30°C, is discussed in more details.

The thesis ends with Chapter 6 - Conclusions, where the main findings and original contributions of this work are given and once more commented.

Samenvatting

Samenvatting behorende bij het proefschrift:
Low-Power High-Accuracy
Smart Temperature Sensors in CMOS Technology

"Slimme" temperatuursensoren met hoge nauwkeurigheid en laag energieverbruik in CMOS technologie door André Luiz Aita

Dit proefschrift behandelt de theorie, het ontwerp en de realisatie van "slimme" CMOS temperatuursensoren met hoge nauwkeurigheid en laag energieverbruik.

Er is een groot aantal temperatuurafhankelijke componenten waarmee temperatuursensoren kunnen worden gemaakt. Als het echter gaat om een volledig geïntegreerde sensor zijn de opties beperkt; in dit geval gebruikt men doorgaans een bipolaire transistor. In CMOS-technologie is al vaak en met succes gebruik gemaakt van de daarin beschikbare compatibel PNP-transistor, maar hiervoor is een fundamenteel begrip van zijn temperatuurafhankelijkheid en zijn niet-idealiteiten essentieel.

Dit proefschrift is als volgt opgebouwd. Na de inleiding wordt in Hoofdstuk 2 - Temperature Sensing Fundamentals - een achtergrond voor het ontwerp van temperatuursensoren geschetst. De eigenschappen en het gedrag van de PNP-transistor worden in detail behandeld, aangezien deze transistor het hart van de temperatuursensor vormt.

Hoofdstuk 3 - Front-end presenteert het ontwerp van een analoge schakeling waarmee nauwkeurig een aantal temperatuurafhankelijke spanningen (zoals de PTAT-spanning ΔV_{BE} en de CTAT-spanning V_{BE}) kan worden gegenereerd. De relatie tussen de instelstromen van twee PNP-transistoren en het energieverbruik en de nauwkeurigheid van de sensor wordt uitvoerig geanalyseerd.

In Hoofdstuk 4 - $\Delta\Sigma$ A/D Converters - wordt het ontwerp van de analoog-digitaalomzetter van de slimme temperatuursensor besproken. Deze omzetter maakt gebruik van de temperatuurafhankelijke spanningen, i.e. ΔV_{BE} and

V_{BE}, om een digitaal signaal te genereren; de waarde van dit digitale signaal is een maat voor de temperatuur. Aan de hand van een A/D-omzetter met een nauwkeurigheid van $\pm 0.1°C$ en een laag energieverbruik wordt een aantal belangrijke afwegingen tussen de nauwkeurigheid en het energieverbruik besproken.

In hoofdstuk 5 - Realizations - worden de meetresultaten van drie versies van een temperatuursensor gepresenteerd en geanalyseerd. De laatste sensor levert de beste prestaties en wordt in meer detail besproken. Deze sensor heeft een onnauwkeurigheid die kleiner is dan $\pm 0.25°C$ ($\pm 3\sigma$) over een temperatuurbereik van 200°C, d.w.z. van $-70°C$ tot 130°C, en heeft een stroomverbruik van $25\mu A$ bij een voedingsspanning van $V_{DD} = 2.5V$ en bij een temperatuur van 30°C.

Het proefschrift wordt afgerond door Hoofdstuk 6 - Conclusions, waarin de voornaamste resultaten en de originele bijdragen van dit werk samengevat en besproken worden.

Sumário

Sumário pertencente à tese:
Low-Power High-Accuracy
Smart Temperature Sensors in CMOS Technology

Sensores de Temperatura Inteligentes de Alta Exatidão e Baixo Consumo em Tecnologia CMOS por André Luiz Aita

Esta tese descreve a teoria fundamental, o projeto e a implementação de sensores de temperatura inteligente de alta exatidão e baixo consumo, em tecnologia CMOS.

Muitos elementos sensíveis à temperatura podem ser usados na construção de sensores de temperatura inteligentes. Contudo, se o sensor é integrado, as opções são mais restritas e o transistor bipolar, em particular o transistor PNP de substrato e compatível com tecnologia CMOS, tem sido freqüentemente utilizado com sucesso. Este sucesso, entretanto, depende do profundo entendimento do seu comportamento com a temperatura e de suas não idealidades para posterior correção.

Esta tese organiza-se como segue.

Após a Introdução, o Capítulo 2 - Fundamentos de Medidas de Temperatura - apresenta os princípios básicos de projeto de sensores de temperatura. As características e comportamento do transistor de substrato PNP são discutidas em detalhe, uma vez que este transistor é o dispositivo fundamental de construção do sensor.

Capítulo 3 - Front-end apresenta o projeto analógico do front-end do sensor, mostrando como gerar com exatidão importantes sinais como a tensão proporcional à temperature absoluta (PTAT) ΔV_{BE} e a tensão inversamente proporcional à temperature absoluta (CTAT) V_{BE}. A correntes de polarização e razões de correntes dos transistores PNP são também analisadas em profundidade em face dos requerimentos de exatidao e limites no consumo de potência.

No Capítulo 4 - Conversores A/D $\Delta\Sigma$, o conversor A/D do sensor de temperature inteligente, i.e. a interface digital do sensor, é analisado e projetado. Este circuito processa a saída analógica do front-end, i.e. os sinais ΔV_{BE} e V_{BE}, gerando uma saída digital, cujo valor é uma representacão digital da temperatura. Importantes relações entre exatidão e consumo de potência são abordados com o foco em um conversor de baixo consumo e inexatidão menor que $\pm 0.1°C$.

No Capítulo 5 - Realizações, resultados de medidas de três versões de um sensor de temperatura são apresentados e analisados. O último sensor, que não é coincidentalmente o sensor com melhor desempenho, apresentando uma inexatidão menor que $\pm 0.25°C$ ($\pm 3\sigma$) sobre a faixa militar estendida de temperatura, i.e. 200°C de $-70°C$ até 130°C, com uma corrente total de $25\mu A$ para $V_{DD} = 2.5V$ e a 30°C, é discutida em mais detalhes.

Esta tese termina com o Capítulo 6 - Conclusões, onde os principais resultados e contribuições originais deste trabalho são retomados e uma vez mais comentados.

Acknowledgments

This thesis is a result of more than 5 years of hard working. And I will not be modest now: its conclusion was possible due to my own efforts. But I have to mention (and I am glad in saying this) that without the professional and non-professional support of many others, my own efforts would not have been enough. I need to thank all people here that somehow have contributed to my work and share my happiness they all have made possible. But I am afraid this will take some pages and I will still miss some names.

Thanks to Prof. Gerard C. M. Meijer for his invitation to come to Delft and his orientation, as my promoter, in many difficult moments. Thanks for sharing with me his amazing experience on temperature sensors but also for providing assistance in several other topics.

Thanks to Prof. Kofi A. A. Makinwa for his critical supervision. The uncountable technical and non-technical discussions have brought me the experience I was looking for.

Thanks to Prof. Johan H. Huijsing for his vital technical orientantion in part of my work. I am amazed with his knowledge and guidance experience.

Thanks to Dr. Michiel A. P. Pertijs for his technical support and supervision during many meetings, helping me to understand some (unexpected) behavior of my sensors, so I could figure out solutions.

Thanks to the head of the department, Prof. Pattrick J. French, who always kept the department up and running.

Thanks to S. Mahdi Kashmiri, for being my colleague, but more importantly, for being my friend, sharing things from Iran that I have learned are not so that different from my country. Thanks for those many barbecue dinners we had. Thanks to him and Esmee for many other dinners. Thanks for his technical help in so many other Ph.D. and life issues. Thanks for hosting me in his home the day before my departure and taking me to the airport.

Thanks to Eduardo Margallo B. and Lorena for the friendship, showing me the best of Spain (including the culinary and wines). Thanks Eduardo for guiding my wife and me through the Camiño de Santiago de Compostela (pilgrimage route to Santiago de Compostela).

Thanks to the brazilian colleague Paulo G. R. Silva (and Janaina), for his valuable tips about Delft and The Netherlands (in a very clear portuguese!). Thanks for his technical support and talks.

Thanks to Chung-Kai Yang for such a pleasant time during the pilgrimage to Santiago and also for making our journey much funnier. Thanks to Sharma Mokkapati for pleasant talks in the corridor but also for hosting me at his place once during a journey to Delft.

Thanks to Gregory Pandraud for his help with french wines, but also for blasting (luckily not completely) my circuits with a laser (which have allowed me to measure some defective samples).

Thanks to Stefan Kostner, my austrian friend for all interesting talks we had while we were learning about belgian and dutch beers. Thanks to him and Babsi for the dinners.

Thanks to Ferran Reverter, my catalan and good friend. Those times supporting the Barcelona team were moments of great relief during my Ph.D. Thanks to him and Laia for the dinners.

Thanks to colleague Caspar van Vroonhoven, for boosting locally (our office) the dutch insights. But also for good talks, technical discussions about everything and for the dutch translations. Thanks to Lukas Mol for bringing some interesting things about vessels and motorcycles, stopping a while the stressing Ph.D. talks.

Thanks to Ali Heidary for motivating me to speak in dutch with him while he was speaking iranian in our table. Thanks to Zili Yu and Rong Wu for showing me some of the chinese culture.

Thanks to Mikhail Loktev (and Rosa) and Oleg Soloviev for the Russian parties that I will never forget. Thanks to the brazilian Davies W. de L. Monteiro, who made my first visit to Delft (for the Ph.D. interviews) much more confortable. And also for introducing me the great russian friends.

Thanks to Frerik Witte and Martijn Snoeij for naturally showing me a bit of the dutch culture. Thanks to Martijn, also, for picking me up (fully loaded) at Delft Central Station. I did not forget.

Thanks to Antoon Frehe, unbelievably always available and willing to solve all unix- and cadence-like bugs. Thanks to Maureen Meekel, for her technical group-support and general help. Thanks for Zu-yao Chang, up stairs, for his laboratory help.

Thanks to the Electronic Instrumentation (EI) Lab. administrative staff, in special to Inge Egmond and Trudie Houweling, for helping me with all matters someone just arrived from abroad has in Delft. Thanks to Willem van der Sluys for his help with all financial matters. Thanks to Joyce Siemers too.

Thanks to my group (Microelectronics Group) at UFSM for remote support, in special my friends Cesar R. Rodrigues and João B. dos S. Martins.

Thanks to CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and UFSM (Federal University of Santa Maria), both in Brazil, as well the EI Lab. of TUDelft for the financial support, without it this Ph.D. would not have been possible.

Thanks to my parents Vinício (in memorian) and Aldonir for their unconditional support during my whole life. You both have thought me the most important things I needed to learn.

Thanks to my sisters, Aline, who took care of my mother while I was in Delft, and Andréa, who showed me a different way of seeing life.

Thanks, from the deep of my hart, to my wife Letícia. who has supported me in many different ways, sharing her life with me, coping with the pitfalls of a long time apart (which have luckily triggered important moments together) and giving us our precious daughter Anna Lara.

Thanks to God for such a long list of names! Thanks to God I have finished this work.

Publications on temperature sensors

M. A. P. Pertijs, A. L. Aita, K. A. A. Makinwa and J. H. Huijsing, *Low-Cost Calibration Techniques for Smart Temperature Sensors.* IEEE Sensors Journal, v.10, n. 6, June 2010, pp. 1098–1105.

A. L. Aita, M. A. P. Pertijs, K. A. A. Makinwa and J. H. Huijsing, *A CMOS smart temperature sensor with a batch-calibrated inaccuracy of* $\pm 0.25°C$ (3σ) *from* $-70°C$ *to* $130°C$, in Proc. IEEE International Solid-State Circuits Conference – Digest of Technical Papers, Feb. 2009, pp. 342–343(343a), San Francisco.

M. A. P. Pertijs, A. L. Aita, K. A. A. Makinwa and J. H. Huijsing, *Voltage calibration of smart temperature sensors*, in Proc. IEEE Sensors, Oct. 2008, pp. 756–759, Lecce.

A. L. Aita and K. A. A. Makinwa, *Low-Power Operation of a Precision CMOS Temperature Sensor based on Substrate PNPs*, in Proc. IEEE Sensors, Oct. 2007, pp. 856–859, Atlanta.

A. L. Aita and K. A. A. Makinwa, *Low-power and accurate operation of a CMOS smart temperature sensor based on bipolar devices and* $\Sigma\Delta$ *A/D converter*, in Proc. IEEE Ph.D. Research in Microelectronics and Electronics Conference, July 2007, pp. 133–136, Bordeaux.

Short biography

André Luiz Aita was born in Santa Maria, RS state, Brazil, in 1968. He received the M.S. in Computer Science – Microelectronics and the Electrical Engineer degrees from Federal University of Rio Grande do Sul – UFRGS, RS, Brazil, in 1995 and 1990 respectively. Since 1997, he is assistant professor at Federal University of Santa Maria, UFSM, Brazil. In the end of 2004, he started pursuing his Ph.D. degree at Delft University of Technology. The topic of his current research is electronic instrumentation, more specifically on low-power and highly-accurate smart temperature sensors in CMOS technology.

André is married with Letícia and has a 22 months old daughter Anna Lara, with whom he currently enjoys his life.

Propositions accompanying the thesis:

Low-Power High-Accuracy Smart Temperature Sensors in CMOS Technology

1. For integrated low-power high-accuracy temperature sensors implemented with standard CMOS technology, the best choice for the temperature-sensing element is the substrate PNP transistor (Chapter 2).

2. In CMOS integrated temperature sensors the temperature-sensing bipolar transistors are usually biased with PTAT currents because they are accurate and easy to generate, and also because they diminish the curvature of the $V_{BE}(T)$. However, PTAT biasing has the drawback that the matching level of the CMOS current-mirrors and the settling time of the circuit are temperature-dependent, which needs to be solved (Chapters 3 and 4).

3. In the design of CMOS integrated temperature sensors, the temperature-sensing bipolar transistors are usually biased with PTAT currents. The error caused by the settling time, which is temperature-dependent because of the PTAT biasing, and the error caused by leakage can significantly be reduced by making the clock frequency PTAT (Chapter 4).

4. The design of integrated smart temperature sensors in standard CMOS technology, with a batch-calibrated (non-trimmed) inaccuracy below ± 0.25 °C ($\pm 3\sigma$) from -70 °C to 130 °C, having a power dissipation less than 65 μW, is feasible (Chapter 5).

5. A PhD project pursued abroad provides a perfect combination of counteracting effects. While the PhD study dominates your mind, forcing it to focus on a very narrow problem, life abroad sets it free to discover a universe of new things that life in a foreign country offers.

6. When living abroad, cultural differences can be successfully managed if you stop trying to understand the other culture and simply accept it as being different from what you know and are used to.

7. Disregard helps when coping with the life ironies we all have to live with, e.g. by not paying attention to countries defending the preservation of the Amazon forest when they have already destroyed their local flora and fauna.

8. Human law has nothing to do with justice, but with the lawyers' ability to persuade in Court.

9. A newborn child is able to add a completely new dimension to his/her parents' world.

10. While administrative rules are necessary to regulate daily situations in any organization, an excess of rules has an effect similar to total deregulation.

André Luiz Aita, Delft, March 1st, 2011.

These propositions are considered opposable and defendable and as such have been approved by the supervisor, Prof.dr.ir. G.C.M. Meijer.

Stellingen behorende bij het proefschrift:

Low-Power High-Accuracy Smart Temperature Sensors in CMOS Technology

1. Voor geïntegreerde laagvermogen temperatuursensoren met een hoge nauwkeurigheid vervaardigd in standaard CMOS technology is voor de keuze van het temperatuurgevoelige element de substraat pnp transistor de beste (Hoofdstuk 2).

2. In CMOS geïntegreerde temperatuursensoren worden de temperatuurgevoelige bipolaire transistoren gewoonlijk ingesteld met PTAT stromen, omdat deze nauwkeurig zijn en gemakkelijk op te wekken, en tevens omdat zij de kromming van $V_{BE}(T)$ curve doen afnemen. De PTAT instelling heeft echter het nadeel dat de ongelijkheid in de CMOS stroomspiegels en de "settling time" van het circuit temperuurafhankelijk zijn, hetgeen een probleem veroorzaakt dat dient te worden opgelost (Hoofdstukken 3 en 4).

3. In het ontwerp van CMOS geïntegreerde temperatuursensoren worden de temperatuurgevoelige bipolaire transistoren gewoonlijk ingesteld met PTAT stromen. De fout die wordt veroorzaakt door de "settling time", welke temperatuurgevoelig is als gevolg van de PTAT instelling, en de fout veroorzaakt door lekstroom kunnen aanzienljk worden gereduceerd door gebruik te maken van een PTAT klokfrequentie (Hoofdstuk 4).

4. Het ontwerp van geïntegreerde temperatuursensoren met een partijgecalibreerde (niet afgeregeld) onnauwkeurigheid van minder dan ±0.25 °C (±3σ) van -70 °C to 130 °C en een dissipatie van minder dan 65 μW in standaard CMOS technologie is haalbaar (Hoofdstuk 5).

5. Een promotieproject dat in het buitenland wordt uitgevoerd voorziet in een perfecte combinatie van twee tegengestelde effecten. Terwijl het promotieproject de geest in beslag neemt, door deze te richten op een zeer smal probleem, wordt deze door het leven in het buitenland bevrijd om de wereld van nieuwe dingen te ontdekken die het leven in een vreemd land te bieden heeft.

6. Wanneer je in het buitenland woont kunnen de culturele verschillen het beste worden gehanteerd door te stoppen met pogingen om de vreemde cultuur te begrijpen en door haar eenvoudig te accepteren als een die verschilt van die waaraan je gewend ben.

7. "Negeren" helpt om om te gaan met de ironieën van het leven, bijvoorbeeld door geen aandacht te besteden aan landen die het behoud van het Amazonewoud verdedigen terwijl zij hun eigen flora en fauna reeds vernietigd hebben.

8. Menselijke wetten hebben niets te doen met gerechtigheid, maar met het vermogen van de advocaten om het hof te overtuigen.

9. Een pasgeboren kind is in staat om een volledig nieuwe dimensie toe te voegen aan de wereld van de ouders.

10. Hoewel voor iedere organisatie administratieve maatregelen nodig zijn om dagelijkse situaties te regelen heeft een overmaat daarvan een effect dat vergelijkbaar is met dat van totale ontregeling.

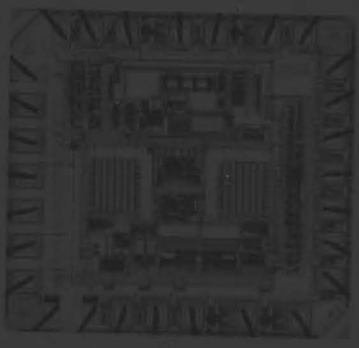
André Luiz Aita, Delft, 1 maart 2011.

Deze stellingen worden opponeerbaar en verdedigbaar geacht en zijn als zodanig goedgekeurd door de promotor Prof.dr.ir. G.C.M. Meijer.

UITNODIGING

Hierbij nodig ik u uit
voor het bijwonen van
de verdediging
van mijn proefschrift:

Low-Power High-Accuracy Smart Temperature Sensors in CMOS Technology



De verdediging zal
plaatsvinden op
dinsdag 1 maart 2011
van 15:00 tot 16:00
in de Senaatzaal
van de Aula van de
Technische Universiteit Delft,
Mekelweg 5 te Delft.

Voorafgaand aan
de verdediging,
zal ik om 14:30 een
korte uitleg geven
over mijn werk.

Na de verdediging bent
u van harte welkom op
de receptie in de
foyer van de Aula.

André Luiz Aita

This thesis discusses the theory, design and implementation of low-power high-accuracy smart temperature sensors in standard CMOS technology. The temperature information extracted using substrate PNP transistors, which are CMOS-compatible, is digitized using an incremental $\Delta\Sigma$ A/D converter. Without trimming, the sensor achieves an inaccuracy of $\pm 0.25°C$ ($\pm 3\sigma$) from -70°C to 130°C. A power consumption of $62\mu W$ is accomplished by optimizing the power/inaccuracy tradeoffs and incorporating several circuit techniques. After a single room-temperature trim, the sensor's inaccuracy is better than $\pm 0.1°C$ ($\pm 3\sigma$).

TU Delft

ISBN 978-90-813316-9-2