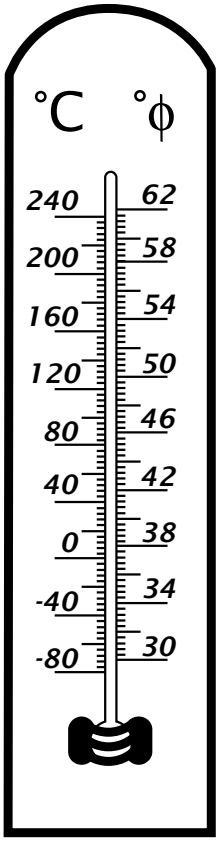
Integrated Temperature Sensors based on Heat Diffusion



Caspar van Vroonhoven

Integrated Temperature Sensors based on Heat Diffusion

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
vrijdag 4 december 2015 om 12:30 uur

door

Caspar Petrus Laurentius VAN VROONHOVEN

Elektrotechnisch Ingenieur,
Technische Universiteit Delft, Nederland
geboren te ‘s-Gravenhage, Nederland

This dissertation has been approved by the
promotor: Prof.dr. K.A.A. Makinwa

Composition of the doctoral committee:

Rector Magnificus	chairman
Prof.dr. K.A.A. Makinwa	promotor, Delft University of Technology

Independent members:

Prof.dr. A.J.P. Theuwissen	Delft University of Technology
Prof.dr. G. de Mey	Ghent University (Belgium)
Prof.dr. B. Nauta	University of Twente
Prof.dr. A.J.M. van Tuijl	University of Twente
Dr. J. Altet	Universitat Politècnica de Catalunya (Spain)
Prof.dr. P.M. Sarro	Delft University of Technology, reserve member

This research was partially funded by National Semiconductor Corporation (now Texas Instruments).

Printed by PrintPartners Ipskamp Drukkers B.V.
ISBN/EAN: 978-94-6259-968-0

Copyright © 2015 by C.P.L. van Vroonhoven

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by an information storage and retrieval system, without written permission of the author.

Opgedragen aan
Nan, Karien en Laurens

Table of Contents

1 Introduction:

Temperature plays a critical role in many physical, chemical and biological processes, and so temperature sensors are widespread. Due to advances in Integrated Circuit (IC) technology, it has become possible to realize *integrated* temperature sensors, small silicon chips that measure temperature and report this information to a controller or an end user [1.1]. Over the last few decades, integrated temperature sensors have become widely used in many consumer, industrial and military applications. This is largely due to their low cost, small size, ease of use, and low inaccuracy.

Integrated temperature sensors are typically based on measuring the temperature-dependent characteristics of bipolar junction transistors (BJTs), which can be implemented in most IC technologies. To further reduce cost, they can be fabricated in relatively mature IC processes, since they do not require the state-of-the-art IC technology found in memory and processor chips.

Historically, integrated temperature sensors have met stand-alone temperature-sensing needs in larger (electronic systems). More recently, however, new applications for integrated temperature sensing have emerged. One of the most important is the thermal management of microprocessors and other large digital chips:

The rapid improvements in CMOS technology have led to exponentially increasing computing power, a process popularly known by “Moore’s Law”. As transistors become smaller, they become faster, and more of them can be placed on the same silicon area: the digital circuit density in microprocessors has thus increased by orders of magnitude. However, the energy density of microprocessors has also gone up significantly, and as such modern microprocessors are now thermally limited, requiring advanced thermal management systems to optimize their performance while maintaining reliability. This requires small, fast and *accurate* temperature sensors.

However, the BJT-based sensors that are so widespread in stand-alone applications exhibit reduced performance when they are implemented in modern CMOS processes. BJTs do not scale down as well as CMOS devices, and the inaccuracy of BJT-based sensors increases dramatically, to several degrees Centigrade. This makes it difficult to meet the needs of thermal management systems in current and future scaled IC technologies.

This thesis will explore a different temperature-sensing principle that should make it possible to realize sensors that are smaller, more accurate and more energy-efficient in modern CMOS technology. Such sensors measure the time it takes for heat to diffuse through silicon. Since phonon transport in silicon is strongly temperature-dependent, the delay in such thermal time-of-flight sensors is a measure of absolute temperature.

The proposed sensing mechanism can be very accurate, since the silicon used in IC fabrication is extremely pure. Additionally, the distance over which the time-of-flight is measured can be more and more accurately defined because of the continuing improvements in optical lithography.

In contrast to BJT-based sensors, the proposed sensor should follow Moore’s Law, and thus offer better temperature-sensing performance in modern CMOS technology. This would provide an effective and future-proof solution to the thermal management of large digital chips. Other applications may also benefit from the proposed sensor, because it uses a fundamentally different sensing principle compared to existing sensors.

This thesis will discuss the modeling and implementation of such thermal time-of-flight sensors. It will also present the circuits needed to read out and digitize their output. Some measurement results will then be presented, followed by conclusions about the viability and usefulness of this new family of sensors.

This chapter will continue with an overview of existing temperature sensors and their applications, including their advantages and disadvantages. Then, the proposed sensor will be positioned in light of the existing prior art and its future potential. The chapter ends with an overview of how this thesis is structured.

1.1 Integrated Temperature Sensing

1.1.1 Existing applications

There are many different applications in which integrated temperature sensors are used. In consumer applications, such as in household appliances and handheld devices, devices typically have to operate over the commercial temperature range, which ranges from 0ºC to +85ºC. Other applications include industrial and automotive markets, ranging from food quality monitoring to climate control, battery management and other process control. Many industrial and military applications require an operating range from -55ºC to +125ºC, or up to +150ºC.

1.1.2 Emerging applications: thermal management in microprocessors

Temperature sensors are increasingly required for the on-chip thermal management of modern microprocessors. Due to their greatly increased energy density, processors have become thermally limited [1.2]. To prevent overheating, the microprocessor's temperature is measured and a cooling mechanism (such as a mechanical fan) is used to limit the die temperature below a certain maximum, T_{max}. This temperature is the maximum allowable by lifetime and reliability constraints. Fig. 1.1 sketches the equivalent thermal system:

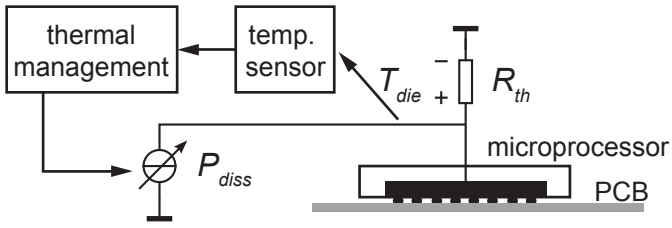


Fig. 1.1: Thermal management in a microprocessor prevents overheating.

This figure shows a basic thermal management system, in which the power dissipated inside the chip, P_{diss}, is regulated as a function of the measured die temperature, T_{die}. Power consumption and processor performance are directly related, so the overall goal is to maximize P_{diss}. This is usually in the order of tens of Watts, so R_{th} is usually designed to be as small as possible, to limit self-heating.

The effectiveness of the above system relies on the accurate measurement of T_{die}. If the temperature sensor overestimates T_{die}, P_{diss} will be reduced unnecessarily. This is exacerbated when R_{th} is low, since an error of a few ºC in the measured T_{die} will then correspond to an error of a few Watts in P_{diss}. On the other hand, if the sensor underestimates T_{die}, temperatures may exceed their safe maximum levels, leading to reliability issues. To prevent this, the thermal management system will usually incorporate a certain temperature guard-band. Having an accurate temperature sensor reduces the width of the guard-band, thus maximizing P_{diss} and improving performance.

The relatively simple system shown above does not meet the requirements of modern microprocessors. These experience extreme energy densities and consist of multiple cores, whose activity depends on the requirements of the operating system and the operational load. Strong thermal gradients can exist across the die, as well as relatively fast thermal transients as cores see varying loads. This leads to thermal *hot spots* that move around the die, and which can be several tens of ºC

hotter than the average chip temperature [1.3]. In such an environment, it is clear that a single sensor cannot effectively manage die temperature.

Modern microprocessors are thus equipped with tens of temperature sensors that are distributed across the die [1.3]. The information provided by these sensors is used to locally throttle the circuitry and thus prevent *local* overheating. Such sensors should be fast enough to sense thermal transients. They should also require as little silicon area as possible, since many sensors are needed and die area is very expensive in advanced CMOS technology. Their inaccuracy should be sufficiently low, and ideally they should require little test effort for calibration and trimming.

Future microprocessors will likely have even more stringent thermal management needs, as they will have more cores and will be realized in denser process nodes. There is thus a need for thermal management systems that follow this trend and are able to ensure reliable operation under all operating conditions.

1.1.3 Emerging applications: wide-range temperature sensing

Extreme-temperature needs from the oil & gas, avionics, automotive and other industries has led to an increasing demand for *integrated* temperature sensors. This is due to the trend towards increasing automation and instrumentation, which is driving requirements on the size, volume and cost of such sensors. For example, in down-hole drilling applications, a complex circuit board including pressure, stress and temperature sensors is mounted close to the drill head, which then goes down the borehole to monitor operating conditions.

Traditionally, such elevated temperatures (e.g. from 175ºC to 300ºC) have been beyond the reach of ICs. This is partially due to reliability concerns such as packaging lifetime, increased drift and more rapid ageing. Despite these challenges, several analog ICs that operate at extreme temperatures have become commercially available in recent years [1.4][1.5]. These include operational amplifiers, analog-to-digital converters, digital-to-analog converters and other analog building blocks. Several vendors already offer specialized high-temperature ICs [1.4][1.5][1.18]. This has been enabled by improved packaging materials and the use of more advanced IC technology, such as silicon-on insulator (SOI) processes.

However, only very few integrated *temperature sensors* are qualified for use at such temperatures, and so even though temperature information is essential in many applications, *discrete* temperature sensors are still widely used. This is largely due to the fact that junction leakage currents increase exponentially with temperature, while their absolute value is relatively poorly controlled. Leakage affects both the sensors and their readout circuits, and so precision circuit design at high temperatures can be very challenging [1.6]. An integrated temperature sensor that *does* operate accurately at these temperatures would offer clear size, reliability and inaccuracy benefits, making their development commercially interesting.

This section has described two emerging applications that are not well served by existing integrated temperature sensors. In order to better understand *why* this is the case, the next section will describe the operating principles and performance of today's mainstream integrated temperature sensors.

1.2 Sensors based on bipolar transistors (BJTs)

1.2.1 Basic principles

In standard CMOS technology, a vertical PNP transistor can be formed; its collector, base and emitter are the substrate, n-well and p^+-diffusions, respectively. This parasitic BJT requires no extra process steps, and is sufficiently insensitive to process spread, bias current levels and mechanical stress [1.7]. As for all BJTs, its base-emitter voltage, V_{BE}, can be approximated as follows:

$$V_{BE} = \frac{kT}{q}\ln(\frac{I_C}{I_S}) \tag{1.1}$$

Due to the strong temperature dependence of I_S, V_{BE} has a temperature coefficient of about -2mV/ºC. The difference in the V_{BE} of BJTs operated at different current densities, ΔV_{BE}, is given by:

$$\Delta V_{BE} = \frac{kT}{q}\ln(p) \tag{1.2},$$

in which p is the ratio between the different current densities. This equation shows that ΔV_{BE} is a linear function of temperature that is essentially independent of process spread and supply voltage. For typical values of p, ΔV_{BE} has a positive temperature coefficient of about 150µV/ºC.

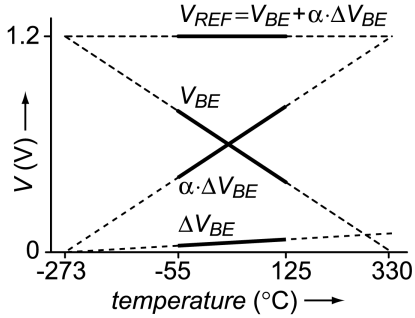


Fig. 1.2: V_{BE} and ΔV_{BE} create PTAT, CTAT and temperature-independent voltages [1.7].

Fig. 1.2 shows how a linear combination of V_{BE} and ΔV_{BE} yields V_{REF}, the well-known bandgap voltage. Since this voltage is nearly temperature-independent, it can act as a reference to V_{PTAT}, so that their ratio is a function of absolute temperature, given by:

$$\mu = \frac{V_{PTAT}}{V_{REF}} = \frac{\alpha\Delta V_{BE}}{V_{BE} + \alpha\Delta V_{BE}} \tag{1.3}$$

Fig. 1.3 shows how both voltages can be extracted and then digitized by an ADC, which can be made to digitally output their ratio µ.

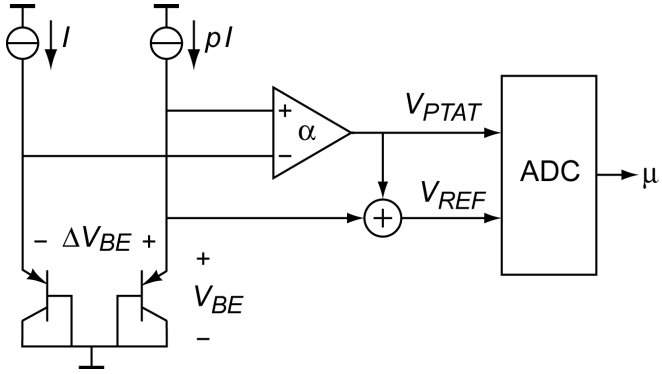


Fig. 1.3: Block diagram of a BJT-based temperature sensor. An ADC outputs the V_{PTAT}/V_{REF} ratio [1.7].

1.2.2 State of the art

BJT-based temperature sensors have been the subject of research for many decades. The first smart temperature sensors were introduced in the 1980's (see e.g. [1.8]). Significant improvements in their accuracy were achieved in the late 1990's and the early 2000's (see e.g. [1.9]), culminating in a design by Pertijs [1.7], which achieved an inaccuracy of ±0.1ºC (3σ) over the military temperature range (55ºC to 125ºC), and represents the state-of-the-art. Fig. 1.4 shows an overview of published results for BJT-based temperature sensors in CMOS technology, based on the data available from [1.28]:

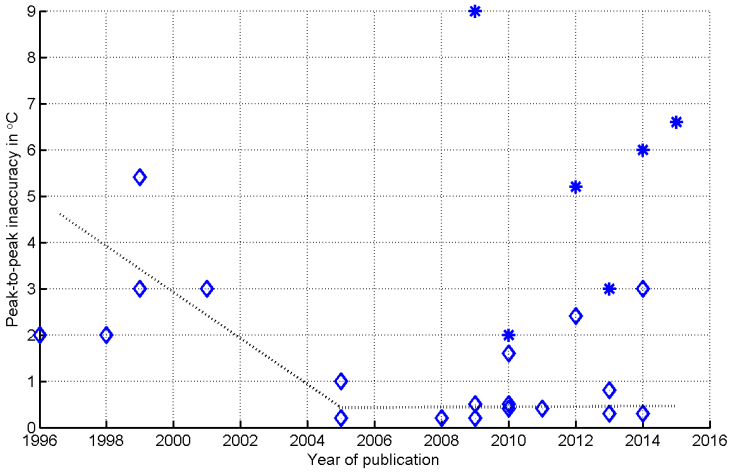


Fig. 1.4: Overview of BJT-based sensor inaccuracy over time based on [1.28].
The data points with a star marker correspond to sensors in deep submicron technologies (< 65nm).

The black dashed line indicates the improvements in inaccuracy over time. The data points marked with a star (*) correspond to sensors in deep submicron technologies (< 65nm). Their inaccuracy is significantly worse, and in the following, the reasons for this will be discussed in more detail.

Recent research has focused on reducing the power consumption of BJT-based sensors. Significant improvements are described in [1.10][1.11][1.12], with the state of the art being a power consumption of 5µW for a resolution of 0.02ºC at a conversion rate of 188S/s [1.12].

As a result of these developments, BJTs have been used in most commercially available smart temperature sensors; e.g. the LM92 [1.13], TMP275 [1.14], ADT7312 [1.15]. However, while BJT-based sensors are quite widespread and well-suited to many applications, they do have several significant limitations.

1.3 Limitations of BJT-based sensors

1.3.1 Trimming

The accuracy of BJT-based temperature sensors is limited by the process spread of I_S. Even when all other error sources are reduced by careful circuit design [1.7], this spread still leads to errors of a few degrees Centigrade. These errors can be reduced by calibrating the sensor against a (more accurate) reference temperature sensor, and then adjusting its output.

Whenever a sensor is compared to a more accurate reference sensor, a *calibration* step takes place. Calibration can be performed at one or multiple temperatures, and results in a calibration table that

describes the sensor's error. Some applications may store this table in the microcontroller that reads out the sensor, so that its error may then be digtially corrected. However, integrated temperature sensors are typically expected to be accurate by themselves (i.e. not require a separate calibration table), and so they are typically programmed to correct for their error. This process is known as *trimming*.

Trimming is costly and time-consuming. The device to be trimmed typically needs to be brought to the target trim temperature, and this requires thermal settling of the packaged sensor, which can take as much as several tens of seconds (depending on package size). Trimming at more than a single temperature point is even more time-consuming and expensive.

Individual trimming, in which each device is individually calibrated and then trimmed, results in the lowest inaccuracy. Inaccuracies as low as ±0.1ºC (3σ) over the military range have been achieved [1.7]. A less costly alternative is *batch* trimming. With this approach, a handful of samples is measured to obtain an average trim code for a large batch of devices (from the same process lot), and this code is then applied to the entire batch. Using this method, inaccuracies of ±0.25ºC (3σ) have been achieved [1.10].

A simpler calibration method is based on the fact that ΔV_{BE} is an accurate function of die temperature, so that a *voltage calibration* can be performed by digitizing ΔV_{BE} with respect to a known (external) reference voltage [1.7] [1.12]. Accuracy levels of ±0.15ºC have been achieved [1.12]. Voltage calibration does not require the package to be at a well-defined temperature, it only requires that its temperature be stable. While this typically requires less time and places less demands on the tester, it is still an extra test step. Moreover, the accuracy of voltage calibration relies upon the robustness of the reverse Early voltage (V_{AR}) to process variations, since this affects the slope of ΔV_{BE} over temperature. For substrate PNPs in mature CMOS processes or for BJTs in BiCMOS technology, V_{AR} can generally be extracted from calibration based on several wafer lots, after which it can be assumed to be sufficiently constant. In deep submicron CMOS technology, little information about the spread on V_{AR} is available in the public domain.

Regardless of which trimming method is used, trimming requires an on-chip programmable memory in which the trim code can be stored and held over lifetime. Especially when many sensors are present on the same die, this can be area-intensive. In high temperature applications, the memories' reliability degrades, and thus loss of trim code (and thus temperature errors) may occur. This can be mitigated by adding redundancy or self-test functionality, but these increase complexity.

A temperature sensor that does not require trimming would cost significantly less, since this will save silicon area and reduce test time, while also improving reliability.

1.3.2 Mechanical stress

Via the piezojunction effect, BJTs are sensitive to mechanical stress [1.16]. Today, most ICs are packaged in plastic packages, which add stress to the die as the encapsulating mold compound cools down. This leads to a 'packaging shift', which is on the order of a few tenths of a degree [1.7]. Additionally, hard "filler" particles in the mold compound may end up close to the BJTs and introduce local stress, which is very difficult to predict.

Trimming before packaging, i.e. at wafer level, is relatively cost-effective. This is because only a single thermal settling step is required to trim thousands of devices. However, if the packaging happens post-trim, the associated stress will result in reduced accuracy. Consequently, wafer-trimmed devices are not as accurate as package-trimmed devices. However, trimming at the package level takes longer and thus costs more.

One way to reduce packaging stress is by coating the die with a pliable stress-reduction layer [1.34], but this may not be enough to completely suppress local stress effects. Additionally, adding such a

layer increases cost and potentially introduces reliability concerns, such as an increased sensitivity to humidity and a limited temperature range.

After an in-package trim, BJT-based sensors will still be sensitive to future changes in packaging stress, for example due to soldering or due to moisture absorption. Clearly, it would be advantageous to find a temperature-sensing principle that is not as sensitive to mechanical stress.

1.3.3 Compatibility with µP thermal management

Parasitic BJTs in CMOS technology do not scale well in deep submicron technology. As shown in Fig. 1.4, their inaccuracy, at several ºC, is rather poor. The best reported untrimmed inaccuracy is ±3ºC [1.35], while the best results with a single trim have an inaccuracy of ±1.5ºC [1.36]. There are three main compatibility issues:

The BJT's current gain, ß, is very low in these technologies, sometimes less than unity. This means that any spread in its nominal value creates relatively large changes in I_C, and hence in V_{BE}. Furthermore, ß is not constant but current-dependent, which makes it more difficult to define accurate collector current ratios, and hence to accurately generate ΔV_{BE}. Although these issues could be solved by using an NPN instead of a PNP, the cost increase associated with the required triple-well process can usually not be justified [1.3].

Although BJTs can decrease in size in deep submicron technologies, they typically scale less well than CMOS devices. BJTs are subsurface junction devices that rely on well rather than on gate geometry; also, the range of current densities over which they can be operated accurately is bounded. In two recent works [1.35][1.36], the sensing core occupies about a 100µm x 100µm area. Especially when many sensors are required, this may add up to a significant cost.

At low temperatures, V_{BE} will be as much as 0.8V, while the supply voltage V_{DD} of modern µPs can be as low as 1.0V; In this case, little headroom is available for the current source that biases the BJT, and channel length modulation can significantly affect the accuracy of the current ratio p. Moreover, future technologies may operate at even lower supply voltages, maybe even lower than V_{BE}, thus exacerbating this problem.

1.3.4 Operation at very high temperatures

Recently, several vendors have released temperature sensors for very high temperature applications. These are intended for automotive and industrial applications, such as down-hole sensing in oil well drills. These sensors are based on BJTs, and their inaccuracy at high temperatures is rather poor: Fig. 1.5 shows the error bounds for two commercially available sensors (as per their datasheet) [1.15][1.17].

The main mechanism behind this increasing error are the strong temperature dependence of I_S and of the leakage currents in the readout circuit, which approximately double for every 10ºC increase in temperature.

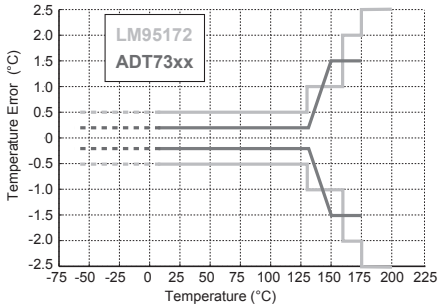


Fig. 1.5: Temperature-sensing inaccuracy of commercially available high T sensors.

Even though recent work [1.19] has shown that, by ensuring that $I_C >> I_S$ even at high temperatures, BJT-based sensors can also be accurate up to 200ºC, their fundamental sensitivity to I_S will make their accurate operation at even higher temperatures an exponentially increasing challenge.

In summary, BJTs are not well suited for the two key applications outlined above. To overcome the fundamental limitations of spread in I_S, minimum voltage headroom requirements and high-T leakage, a different temperature-sensing principle is required. Ideally, this principle should not require trimming, and its performance in modern CMOS technologies should be at least as good as in older ones. Section 1.5, and in fact the rest of this thesis, will propose such a mechanism. Before doing so, however, this thesis will first briefly discuss the existing alternatives to BJT-based sensors.

1.4 Existing alternatives to BJT-based sensors

Although BJTs are by far the most widely used temperature sensing elements in CMOS technology, there have also been some research efforts aimed at using MOSFETs or resistors as temperature-sensing elements. These will be discussed very briefly here; for more information, the reader is referred to [1.28] and [1.29].

1.4.1 MOSFET-based sensors

In a MOSFET, the relationship between gate voltage and drain current is temperature-dependent. When a MOSFET is biased in weak-inversion, this relationship is very similar to the V_{BE}-I_C relationship in a BJT, so that BJTs can simply be replaced by MOSFETs in circuits such as the one shown in Fig. 1.3. Alternatively, MOSFETs can be used to create ring oscillators or delay lines, of which the output frequency or the delay, respectively, is temperature-dependent.

Unfortunately, the untrimmed inaccuracy of MOSFET-based sensors is quite poor. This is because both their threshold voltage V_{TH} and channel mobility μ suffer from process spread, and thus the MOSFET's temperature characteristic has at least two degrees of freedom. The state of the art inaccuracy is ±0.4ºC (3σ) after a single-point trim [1.30], but most publications report an inaccuracy of between ±0.5ºC and ±2.0ºC after a two-point trim [1.28]. The MOSFET is sensitive to leakage currents in a way similar to BJTs.

In the context of deep-submicron temperature sensing, one of the advantages of using a MOSFET is their lower headroom requirement, since the V_{TH} of modern MOSFETs can be significantly lower than V_{BE}. MOSFET-based sensors scale well, and are typically very well-modeled. However, the need for trimming still makes them unattractive for many-sensor, multi-core thermal management applications.

1.4.2 Resistor-based sensors

Most resistors available in CMOS processes are temperature-dependent, thus providing another mechanism of sensing die temperature. One of the key advantages of using resistors is that they are typically strongly temperature-dependent and can be designed for low noise, and thus are able to achieve a very high resolution (sub-mK) [1.31][1.32].

In the context of deep submicron temperature-sensing, their advantages are similar to those of MOSFETs. Resistors can be operated under arbitrarily low headroom, and they scale reasonably well. However, resistor-based sensors are quite inaccurate. Their resistance depends on geometry as well as on doping concentration, making it difficult to achieve good untrimmed results. Moreover, it is non-trivial to implement an on-chip (temperature-independent) reference resistance. The lowest reported inaccuracy is ±0.35ºC after a one-point trim [1.33].

Although both MOSFET-based and resistor-based temperature sensors have merit, they suffer from the same trimming problem as BJT-based sensors. The next section of this thesis will explore another principle that may allow for *untrimmed* temperature sensors: thermal diffusivity sensing.

1.5 Proposed solution: thermal diffusivity sensing

1.5.1 Electrothermal Filters

The thermal diffusivity of silicon, D_{Si}, determines the time it takes for heat to diffuse through silicon, in this case the substrate of an IC. D_{Si} is proportional to $1/T^{1.8}$, mainly due to the strong temperature dependence of phonon scattering in crystalline semiconductors. (The same scattering mechanism also contributes to the temperature dependence of the *electrical* mobility, μ.)

Because of the temperature dependence of D_{Si}, the thermal delay between a heater and a relative sensor, spaced at distance r, will also be temperature-dependent. Such a structure is known as an Electrothermal Filter, or ETF. It is schematically shown in Fig. 1.6:

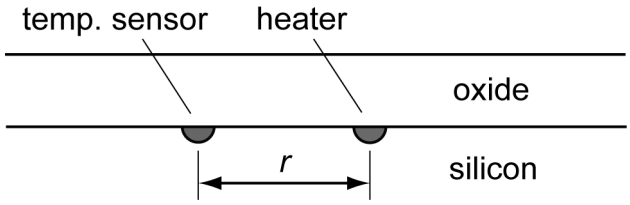


Fig. 1.6: A basic electrothermal filter (ETF).

ETFs have two important accuracy advantages. Firstly, D_{Si} is essentially process-independent, because the silicon used in IC technology is highly pure. It is also only very weakly sensitive to doping fluctuations [1.20]. Secondly, the distance r can be very accurately defined by lithography.

For a spacing of 20µm, the typical thermal delay at room temperature is about 2.5µs. Compared to typical electrical delays, heat diffusion is thus a relatively slow process. This suggests that it should be possible to make circuits that measure this delay with sufficient accuracy.

A temperature sensor based on the above ETF should measure its thermal delay and compare it to a (temperature-independent) reference delay, which can be provided by a crystal oscillator. From their ratio, absolute temperature can be calculated.

It will be clear that, to first order, such sensors should not suffer from any of the drawbacks of BJT-based sensors. Their inaccuracy can be expected to improve with CMOS scaling, because r will be

more and more accurately defined. ETFs are unaffected by leakage, since they operate in the time domain. They may thus be an interesting alterative to BJT-based sensors.

One of the main drawbacks of ETF-based temperature sensing is that an ETF's heater dissipates a significant amount of power. The sensor's resolution is directly proportional to the power dissipated in the heater, and even at mW-level power dissipation, it tends to be several orders of magnitude lower than that of BJT-based sensors. However, as will be shown later on in this thesis, this metric can be improved by taking advantage of the increasing lithographic resolution to reduce the distance between heater and sensor and scale down the ETF.

As mentioned above, the ETF needs a reference delay to measure absolute temperature. If an accurate clock is *not* available externally, it should be co-integrated with the sensor. However, most on-chip time references suffer from poor inaccuracy, and using those would negate the ETF's low inaccuracy.

A possible solution to the time reference problem relies on sensing the nearly temperature-independent thermal diffusivity of amorphous (rather than crystalline) solids. In IC technology, there are several high-quality amorphous silicon dioxide (SiO_2) structures. Combining such an 'oxide' ETF with a 'silicon' ETF (Fig. 1.6) may lead to a self-referenced temperature sensor.

There is a substantial amount of prior art on ETFs and other thermal-domain devices. Before moving on to present more details about thermal-diffusivity-based temperature sensors, this prior art will first be reviewed.

1.5.2 Early History

Ever since the early days of ICs, the thermal-domain behavior of integrated circuits has been a subject of considerable interest. Historically, self-heating has always required an understanding of the thermal resistance and heat capacity of ICs. Moreover, due to initial difficulties in implementing long time constants with on-chip electrical circuits, the "slower" thermal-domain was considered as a viable alternative. In early works published in the 1960's and 1970's [1.21][1.22][1.23], the use of ETFs in low-pass and band-pass filters with corner frequencies in the 10Hz – 10kHz range was explored. Their use as the time-constant-defining elements of oscillators was also suggested [1.24].

However, thermal-domain circuits did not develop much further. This was most likely due to the rather complicated underlying physics, the relatively high power consumption and the availability of other methods. For example, it became much more straightforward to implement large time constants with the introduction of switched-capacitor circuits, which were far less area- and power-intensive.

1.5.3 Temperature sensing

Probably the first work in which the thermal diffusivity of silicon was used for on-chip *temperature sensing* was that of Szekely et al. [1.25]. Their main idea was to use the thermal delay as the basis of a relaxation oscillator, whose output frequency would then be temperature dependent. While this system clearly demonstrated the soundness of the basic concept, the limited lithographic accuracy and readout circuit development of the time led to poor performance.

Since then, however, CMOS technology has greatly improved: deep-submicron lithographic accuracy and ppm-level crystalline purity, combined with many advances in both digital and analog signal processing now allow the realization of advanced thermal diffusivity sensing structures and high resolution readout architectures. As we shall see shortly, these developments reduce or remove the original limitations on thermal diffusivity sensors, and thus a second look at thermal diffusivity-based temperature sensors is warranted.

1.6 A second look at thermal-diffusivity based temp sensors

Recent work in 0.7μm CMOS technology has shown that TD sensors with improved accuracy and reasonable power efficiency can be realized with the help of new readout architectures [1.26][1.27]. These will be discussed in more detail in Chapters 2 and 3. Based on these results and from first principles, further ETF development holds several promises:

1.6.1 Promises

- As argued above, the main source of sensor inaccuracy is lithographic inaccuracy. This is one of the strengths of CMOS technology: at every new process nodes, lithography becomes more accurate. ETFs should therefore also become more accurate, or they can become smaller and more power efficient.
- Even in older technologies, D_{Si} and r should be sufficiently well-defined to enable temperature sensors that do not require *any* trimming. This is a great step forward for integrated temperature sensors, offering significant cost advantages.
- Modern readout circuits can be made very fast, so that any non-thermal delay will only cause a very small corresponding temperature error.
- Since ETFs operate in the time domain, they should be insensitive to the leakage currents that plague other sensors at extremely high temperature. Therefore, they may be able to service the very high temperature sensing applications.

Clearly, the first three items tie in well with the need for better thermal management sensors, while the last item holds promise for extreme range temperature sensing.

1.6.2 Challenges

- Although recent work has already shown significant improvements in the power efficiency of ETFs, at several mW's, they are still orders of magnitude more energy intensive than their BJT-based counterparts. This will have to be improved by increasing sensitivity, better sensor design and technology scaling.
- Little is known about the ETF's performance as a function of process technology. Although first principles show that ETFs should scale, this will still need to be proven.
- To measure absolute temperature using a single ETF's thermal delay, a reference clock is required. While μPs typically operate from a crystal oscillator, some other applications may not have such a clock available. The search for a sufficiently accurate delay reference is an important challenge in this work. One promising approach to this problem is the use of a second ETF with its thermal properties based mostly on silicon dioxide instead of silicon.
- Even if ETFs live up to their expectations, they will have to clearly outperform the well-established BJT-based sensors and sensor architectures in order to actually replace them.

1.7 Research questions and thesis organization

The main research question of this thesis is this:

Can Electrothermal Filters be used as competitive integrated temperature sensors?

This question leads to three sub-questions, which are as follows:

1. ***How should ETFs be modeled?***
2. ***How should ETFs be read out?***
3. ***How do ETFs perform?***

The answers to these sub-questions should all suggest a positive answer to the main question, or at least their answers should not preclude a future positive answer.

Based on the promises and challenges discussed above, it is likely that ETFs will be most competitive for thermal management in deep submicron CMOS technology, and potentially for wide-range temperature sensing, due to their insensitivity to leakage.

To be competitive in thermal management, it is important to show the scalability of ETF-based temperature sensors; in other words, their accuracy, power consumption and layout area should all improve along the trend of Moore's law. Since most microprocessors have an accurate time reference available, only a sensing ETF is required.

To be competitive in high-temperature applications, ETF-based sensors should be accurate over a wide temperature range. Importantly, in such applications, an accurate time reference may not always be available, so that this application requires a fully self-referenced sensor; thus requiring the use of a sensing and a reference ETF.

The chapters of this thesis are structured as follows:

- This chapter introduces integrated temperature sensing and includes a brief overview of today's mainstream sensors. It discusses their operating principles, their advantages and disadvantages, and important applications and markets. It also shows that there is room for improvement, and how thermal diffusivity-based sensors can compete with existing integrated sensors.
- Chapter 2 discusses the physical background, theory and modeling of ETFs in an attempt to answer sub-question one. It starts by providing some basic background information on the thermal properties of solids, in particular semiconductors. It then shows how thermal diffusivity of silicon and silicon dioxide can be measured using ETFs, and how ETFs can be designed for optimal performance. The chapter finishes with a number of sections that each discuss the modeling and scalability of one aspect of ETFs.
- Chapter 3 discusses ETF systems, which are the systems that can be built around ETFs to leverage their accurate thermal delay characteristics. These include single-ETF temperature sensors, frequency references and self-referenced (multi-ETF) temperature sensors. System-level requirements on accuracy, resolution and power consumption will be translated into specifications for a multi-purpose ETF readout circuit.
- Chapter 4 then discusses this readout circuit in more detail. A phase-domain sigma-delta modulator is used to convert the ETF's phase shift, ϕ_{ETF}, to a digital value by comparing it to a reference phase shift. An implementation in a 0.18µm process is discussed in detail, and the chapter finishes by showing measurements results for a standalone PDΣΔM. Together with chapter 3, this chapter attempts to answer sub-question two.
- Chapter 5 presents measurements results for several test devices and provides answers to sub-question three. The theoretical predictions launched in chapter 2 are confirmed or deviations are explained where necessary. The reader looking for an overview of measured accuracy, resolution and scaling properties of ETFs is referred here.
- Chapter 6 attempts to answer the main research question and summarizes the main findings of this thesis. It also discusses possible future work for ETF-based temperature sensing and other research venues.

1.8 References

[1.1] A. Bakker, “CMOS Smart Temperature Sensors – An Overview”, *Proceedings of IEEE Sensors*, pp. 1423-1427, 2002.

[1.2] H. F. Hamann et al., "Hotspot-Limited Microprocessors: Direct Temperature and Power Distribution Measurement," *IEEE J. Solid-State Circuits*, vol. 42, no. 61, pp. 56-65, Jun. 2007.

[1.3] Y.W. Li and H. Lakdawala, "Smart integrated temperature sensor - mixed-signal circuits and systems in 32-nm and beyond," *IEEE Custom Integrated Circuits Conference (CICC),* pp.1-8, Sept. 2011

[1.4] High Temperature Standard Products, CISSOID, available online through www.cissoid.com

[1.5] Texas Instruments Applications: Down Hole Drilling, Texas Instruments, available online through http://www.ti.com/solution/down_hole_drilling

[1.6] P. C. de Jong, G. C. M.Meijer, and A. H. M.van Roermund, "A 300°C dynamic-feedback instrumentation amplifier," *IEEE Journal of Solid State Circuits*, vol. 33, pp. 1999-2009, Dec. 1998.

[1.7] M. A. P. Pertijs and J.H. Huijsing, *Precision Temperature Sensors in CMOS Technology*, Springer, 2006.

[1.8] 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.

[1.9] A. Bakker and J.H. Huijsing, *High-Accuracy CMOS Smart Temperature Sensors.* Boston: Kluwer Academic Publishers, 2000.

[1.10] A. L. Aita, “Low-Power High-Accuracy Smart Temperature Sensors in CMOS Technology”, Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, Mar. 2011.

[1.11] K. Souri, K.A.A. Makinwa, “A 0.12 mm2 7.5 µW micropower temperature sensor with an inaccuracy of ±0.2 °C (3σ) from –30°C to 125°C”, *IEEE Journal of Solid State Circuits*, vol. 46, no. 7, pp.1693-1700, July 2011.

[1.12] K. Souri, Y. Chae, K.A.A. Makinwa, "A CMOS Temperature Sensor With a Voltage-Calibrated Inaccuracy of ±0.15ºC (3σ) from 55 ºC to 125 ºC," *Solid-State Circuits, IEEE Journal of* , vol.48, no.1, pp.292,301, Jan. 2013

[1.13] LM92 datasheet, Texas Instruments, available online through http://www.ti.com

[1.14] TMP275 datasheet, Texas Instruments, available online through http://www.ti.com

[1.15] ADT7312 specifications, Analog Devices, available online through http://www.analog.com

[1.16] J. Creemer, F. Fruett, G. Meijer, and P. French, “The piezojunction effect in silicon sensors and circuits and its relation to piezoresistance,” *IEEE Sensors J.*, vol. 1, no. 2, pp. 98 –108, Aug. 2001.

[1.17] LM95172 datasheet, Texas Instruments, available online through http://www.ti.com

[1.18] X-Grade products, Linear Technology, available online through www.linear.com/parametric/X_Grade_Products

[1.19] K. Souri, K. Souri, K. Makinwa, "A 40µW CMOS temperature sensor with an inaccuracy of ±0.4°C (3σ) from −55°C to 200°C," *ESSCIRC 2013,* pp. 221-224, Sept. 2013.

[1.20] M. Asheghi et al., “Thermal conduction in doped single-crystal silicon films”, *Journal of Applied Physics*, Vol. 91, pp. 5079-5088, April 2002

[1.21] P. R. Gray and D. J. Hamilton, “Analysis of electrothermal integrated circuits,” *IEEE J. Solid-State Circuits*, vol. 6, no. 1, pp. 8–14, Feb. 1971.

[1.22] W. J. Louw, D. J. Hamilton, and W. J. Kerwin, “Inductor-less, capacitor-less electrothermal filters,” *IEEE J. Solid-State Circuits*, vol. 12, pp. 416–423, 1977.

[1.23] W. T. Matzen, R. A. Meadows, J. D. Merryman, and S. P. Emmons, “Thermal techniques as applied to functional electronic blocks,” *Proc. IEEE*, vol. 52, pp. 1496–1501, Dec. 1964.

[1.24] G. Bosch, “A thermal oscillator using the thermo-electric (seebeck) effect in silicon,” *Solid-State Electron.*, vol. 15, no. 8, pp. 849–852, 1972.

[1.25] V. Szekely, “Thermal monitoring of microelectronic structures”, *Microelectronics J.*, vol. 25, no. 3, pp. 157–170, 1994.

[1.26] K.A.A. Makinwa and J.F. Witte, “A Temperature Sensor based on a Thermal Oscillator,” *Proc. IEEE Sensors*, pp. 1149 – 1152, Oct. 2005

[1.27] K.A.A. Makinwa and M.F. Snoeij, “A CMOS temperature-to-frequency converter with an inaccuracy of ±0.5°C (3s) from –40 to 105°C,” *J. Solid-State Circuits*, vol. 41, is. 12, pp. 2992 – 2997, Dec. 2006.

[1.28] K.A.A. Makinwa, "Smart Temperature Sensor Survey", [Online]. Available: http://ei.ewi.tudelft.nl/docs/TSensor_survey.xls

[1.29] K.A.A. Makinwa, “Smart temperature sensors in standard CMOS”, *Proc. Eurosensors XXIV*, pp. 930-939, Sep. 2010.

[1.30] K. Souri et al., “A 0.85V, 600nW All-CMOS Temperature Sensor with an Inaccuracy of ±0.4°C (3σ) from -40°C to 125°C”, *ISSCC Digest*, pp. 222-223, Feb. 2014.

[1.31] Shahmohammadi, M. et al., "A resistor-based temperature sensor for MEMS frequency references," *Proc. ESSCIRC 2013,* pp.225-228, Sept. 2013.

[1.32] Perrott, M.H. et al., "A Temperature-to-Digital Converter for a MEMS-Based Programmable Oscillator With Frequency Stability and <1ps Integrated Jitter," *J. Solid-State Circuits*, vol.48, no.1, pp.276-291, Jan. 2013.

[1.33] Chan-Hsiang Weng et al., "A CMOS thermistor-embedded continuous-time delta-sigma temperature sensor with a resolution of 0.01°C," *Proc. IEEE ASSCC*, pp.149-152, Nov. 2014.

[1.34] G.Ge et al.,”A Single-Trim CMOS Bandgap Reference With a 3 σ Inaccuracy of ± 0.15% From 40ºC to 125 ºC," *J. Solid-State Circuits*, vol.46, pp.2693-2701, Nov. 2011.

[1.35] R. Eitan and A. Cohen, “Untrimmed Low-Power Thermal Sensor for SoC in 22 nm Digital Fabrication Technology”, *J. Low Power Electronics and Applications,* vol. 4, pp 304-316, Dec. 2014.

[1.36] J.S. Shor and K. Luria, "Miniaturized BJT-Based Thermal Sensor for Microprocessors in 32- and 22-nm Technologies," *J. Solid-State Circuits*, vol. 48, no. 11, pp. 2860-2867, Nov. 2013.

2 Electrothermal Filters

From physics to implementation

2.1 Introduction

This chapter will discuss how the thermal diffusivity of silicon D_{Si} and silicon dioxide D_{SiO2} can be measured using electrothermal filters (ETFs). Such filters are essential components of thermal-diffusivity-based temperature sensors, and thus understanding their operation is a first step towards answering the main research questions of this thesis.

First, the physical origin and temperature dependence of several key thermal properties of Si and SiO_2 will be discussed. It will then be shown how ETFs can be used to measure D_{Si} and D_{SiO2}, and how they can be implemented in CMOS technology. Lastly, various electro-thermal behavioral models of ETFs are developed and used to estimate ETF performance (e.g. accuracy, operating range and power consumption) in several practical CMOS process technologies.

2.2 Thermal properties of Silicon and Silicon Dioxide

2.2.1 Introduction

To understand how ETFs work and how they behave over temperature, some background information on the thermal properties of silicon and silicon dioxide is required. For all structures in this work, the effects of convection and radiation are negligible, and thus only the effect of thermal conduction will be considered.

The two properties of materials most relevant to this work are their *ability to store heat* and their *ability to transport heat*. According to classical thermodynamics, these abilities are quantified by the material's heat capacity C_v and thermal conductivity k, respectively. The ratio k/C_v then determines the rate at which heat diffuses through a material and is known as its thermal diffusivity D.

While these material properties are usually explained in terms of classical thermodynamics, they can also be understood in terms of quantum physics. Although a comprehensive overview is beyond the scope of this work [2.4][2.5], an intuitive summary of some of the key concepts will lead to a better understanding of ETF behavior.

In crystalline solids, atoms (or molecules) are arranged in an ordered lattice. At zero Kelvin, the lattice is in thermal equilibrium and the atoms are static. As the temperature increases, the atoms start to move around their equilibrium position, causing the lattice to vibrate [2.5]. These lattice vibrations, or "the concerted harmonic motion of all atoms" [2.1], store thermal energy, and the amount of stored energy determines the material's heat capacity.

Because atoms in a solid are closely coupled, so are their vibrations. When the lattice is locally heated, the associated increase in vibrational energy will be transported to other parts in a wave-like manner. In electrically non-conducting solids, the speed and the mean free path of these waves determine the thermal conductivity. In (semi-) conductors, free electrons also contribute to the thermal conductivity (as is evident from the fact that metals tend to be good conductors of heat). For the semiconductors that are typically used in IC technology, however, the thermal conductivity is mainly determined by lattice vibrations.

In a way similar to for electromagnetic waves, a quantum-mechanical wave-particle duality exists. There are several modes in which atoms or sets of atoms can vibrate. The energy for each of these modes is quantized into a discrete number of *phonons* [2.2]. Analogous to charge carriers like

electrons and holes (but having a quite different physical origin), phonons can be considered to be *heat carriers*. This analogy is helpful in developing an intuitive understanding of ETFs.

Increasing temperature increases the number of phonons in a lattice, and temperature differences in the lattice cause phonon transport. The heat capacity and thermal conductivity associated with these two mechanisms are discussed in more detail below.

2.2.2 Heat capacity

Heat capacity is a measure of the ability of a material to store heat. Consider an isolated bar of a crystalline solid with length L and surface area A, as shown in Fig. 2.1. Classical thermodynamics [2.5] states that when an amount of heat (Q, in Watt) is added to this volume ($V=LA$), the temperature rise of this bar (ΔT, in Kelvin) is given by

$$\Delta T = \frac{Q}{C_V V}, \tag{2.1}$$

in which C_V is the volumetric heat capacity (in J/cm^3/K). For crystalline silicon at room temperature, $C_V = \rho C_p$ is about 1.64J/cm^3/K, while for silicon dioxide, C_V is about 1 J/cm^3/K.

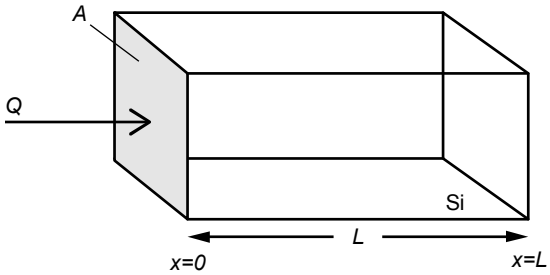


Fig. 2.1: Heat flowing into a volume of Silicon.

The heat capacity of solids is temperature-dependent, and this dependence is best understood with the help of the phonon model discussed above. At zero Kelvin, there are no phonons, i.e. the lattice does not vibrate and therefore cannot store heat. As temperature increases, more phonons are generated and, to store the increase in their vibrational energy, the wavelength of the highest-frequency phonon in the lattice gradually decreases. At, and beyond, the Debye Temperature, T_D, the lattice has sufficient energy to allow all possible modes of oscillation; as a result C_v does not increase further and approaches the Dulong-Petit law, which states that the molar heat capacity of a solid converges to $3R$, with R being the universal gas constant (R=8.31J/mol/K).

The Debye model approximates the temperature dependence of C_v as a function of (T/T_D). For silicon, $T_{D,Si}$ = 645K, while silicon dioxide has a lower Debye temperature ($T_{D,SiO2}$ = 290K). According to the Debye model $C_V(T)$ is given by [2.2][2.5]:

$$C_V(T) = 9Nk_b\left(\frac{T}{T_D}\right)^3 \int_0^{T_D/T} \frac{x^4 e^x}{(e^x-1)^2} dx\,, \tag{2.2}$$

in which N is the number of atoms per cm^3 and k_b is Boltzmann's constant. The unitless integration variable x represents the fraction of the energy levels of the phonon oscillation modes that are excited, being equal to one at T_D (since all modes are excited at T_D). Eq. (2.2) can be evaluated numerically to yield $C_{V,Si}$ and $C_{V,SiO2}$ as a function of temperature from 0K to T_D:

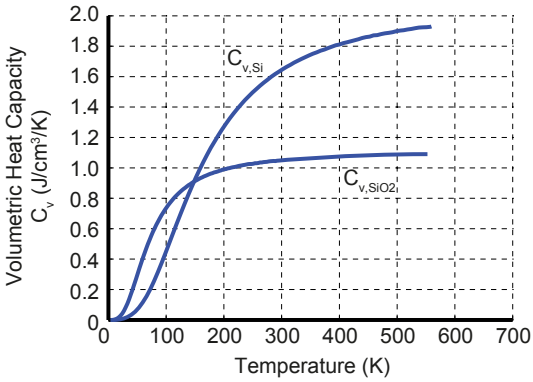


Fig. 2.2: Heat capacity of Si and SiO_2 as a function of temperature.

As can be seen from this figure, $C_{v,Si}$ changes significantly over the military temperature range (218 – 398K), while $C_{v,SiO2}$ is nearly temperature-independent over this range (owing to its lower T_D). In this model, the value for N has been kept constant, since the temperature dependence of the density of silicon is very small (~2ppm/°C at room temperature). Interestingly, the heat capacity of SiO_2 is one of the few IC properties that is nearly constant over the military temperature range.

Fig. 2.2 also shows that $C_{v,Si}$ and $C_{v,SiO2}$ have comparable magnitudes. This has the important consequence that structures in CMOS technology have an effective heat capacity that is mostly determined by $C_{v,Si}$, since silicon structures (e.g. epitaxial and bulk layers) are typically orders of magnitude more voluminous than SiO_2 structures (such as gate and field oxides).

As will be discussed later in this chapter, it is instructive to note the analogies between the thermal and the electrical domain – C_V (J/K) is analogous to the electrical capacitance (Q/V).

2.2.3 Thermal Conductivity

2.2.3.1 General concept

In its most general form, the thermal conductivity of a solid relates the heat flow through it to the temperature gradient across it. When the Fig. 2.1 structure is no longer fully isolated but rather is kept at a constant temperature at $x=L$, a constant flow of heat entering the structure at x=*0* will cause a constant temperature difference ΔT (in Kelvin) across the bar. Fourier's law states that ΔT is given by [2.3]:

$$\Delta T = \frac{l}{kA} Q \tag{2.3}$$

In this equation, k is the thermal conductivity in W/m/K. Again, an analogy with the electrical domain can be made, and the electrical equivalent of Eq. 2.3 is Ohm's law: k^{-1} is a thermal resistivity and the thermal resistance of the structure in Fig. 2.1 is defined as:

$$R_{th} = \frac{l}{kA} \tag{2.4}$$

Because k is finite, R_{th} is finite and systems that dissipate power tend to self-heat.

In IC technology, R_{th} is best known for its role in calculating the self-heating of chips. Its value is a function of die and package characteristics that is usually empirically established. However, the scale on which R_{th} is relevant for ETFs is much smaller (i.e. at the transistor level), and determining its

value requires a deeper understanding of heat conduction and its temperature dependence. Again, the phonon model discussed in section 2.2.1 will provide useful.

The thermal conductivity of solids is determined by (largely independent) contributions from phonon- and free electron heat transport [2.1], so that:

$$k_{tot} = k_e + k_{ph}$$

In metals, k is mostly determined by the large number of free electrons that contribute to heat conduction so that $k_e >> k_{ph}$. It is well known that metals generally have a high thermal conductivity (e.g. 400W/m/K for Copper, 240W/m/K for Aluminum). Because k is dominated by electron heat transport, it is related to the *electrical* conductivity through the Wiedemann-Franz law,

$$k\rho \propto T \ , \tag{2.5}$$

which states that the product of the thermal conductivity and the electrical resistivity is proportional to temperature. For most metals, ρ has an approximately linear temperature dependence, so that k is approximately constant over temperature.

In non-metallic crystalline solids such as silicon, k_e is small and heat conduction is instead dominated by transport through lattice vibrations (i.e. through phonons). As discussed earlier in this chapter, when an increase in temperature in one part of a lattice causes a local increase in vibrational energy, this energy will be distributed through the lattice by phonon transport.

Modern thermal conductivity theory models heat transfer as a phonon transport process, in which phonons carry heat while scattering through the lattice at a certain average velocity, v_s, and having a mean free path λ_{ph}. The thermal conductivity can then be expressed by [2.1]:

$$k = \tfrac{1}{3} C_V v_s \lambda_{ph} \tag{2.6}$$

As this equation shows, the analogy with electrical conductivity still holds: C_v represents the number of phonons available, and $v_s\lambda_{ph}$ represents their mobility.

The mean phonon velocity, v_s, is approximately equal to the speed of sound in the crystal [2.2], and is about 8433m/s for silicon.

The mean free path, λ_{ph}, defines the mean distance traveled by phonons between scattering events. There are several phonon scattering mechanisms, each having strongly different behavior over temperature.

For bulk silicon, the three most relevant scattering mechanisms are [2.1]:

1. Phonon-boundary scattering. At low thermal energy (low temperature), there are few phonons and the energy associated with their vibrations is small. The mean free path λ_{ph} increases exponentially as temperature decreases, until it is limited by the crystal boundaries. For mono-crystalline silicon, λ_{ph} increases up to the sample size; for poly-crystalline silicon, scattering on grain boundaries significantly reduces λ_{ph}.
2. Phonon-impurity scattering. Phonon transfer is most efficient in a perfect lattice structure. When impurities such as dopant atoms or interstitial defects exist, phonons can scatter on these lattice defects. Equivalently, the associated mass differences cause the coupled harmonic oscillators to be less efficient.
3. Phonon-phonon scattering. At the relatively high temperatures of the military temperature range (-55ºC to 125ºC), many phonons exist and have sufficient energy so that phonon-

phonon interactions limit λ_{ph}. For all cases and operating regimes described in this work, phonon-phonon scattering is dominant.

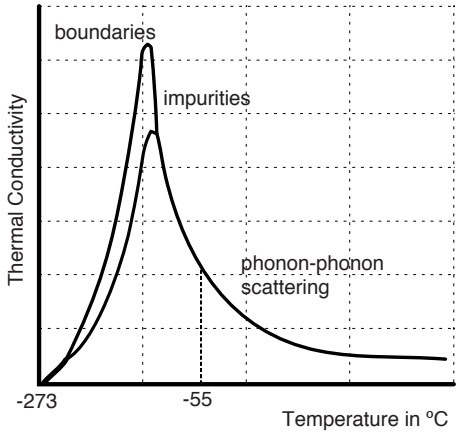


Fig. 2.3. Thermal conductivity as a function of temperature; curves reproduced from [2.1]. The curve with higher k corresponds to a larger sample size.

The temperature dependence of k can be understood by evaluating the combined temperature dependence of Eq. (2.6). C_v was discussed in section 2.2.2 and v_s is nearly temperature-independent [2.2], but λ_{ph} is a strong function of temperature. Fig. 2.3 shows a qualitative overview of k_{Si} as a function of absolute temperature, also indicating which phonon scattering mechanisms limit k_{Si}.

2.2.3.2 Silicon

Although the use of mean values makes Eq. (2.6) relatively simple, a much more complex analysis of all the different types of phonons and their scattering rates is required for the accurate modeling of k_{Si}. The work of Holland [2.6] provided a model for calculating k_{Si} as a function of temperature for bulk silicon. The model was later expanded by Asheghi et al. [2.7] to also incorporate the effect of doping and finite layer thickness, by adding terms to λ_{ph} to incorporate phonon-impurity and phonon-boundary scattering, respectively. Using these models, k_{Si} for undoped silicon was calculated as shown in Fig. 2.4. The temperature range was limited to -70ºC to 250ºC, which covers all the applications discussed in this work.

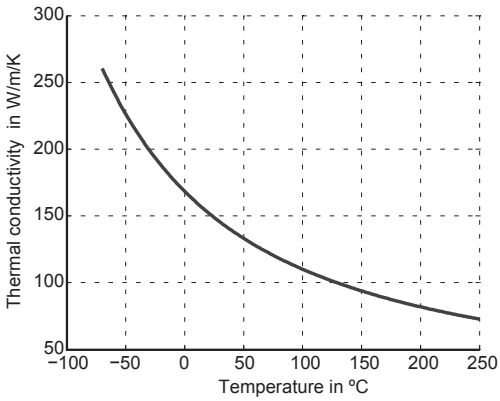


Fig. 2.4: Silicon thermal conductivity as a function of temperature.

The resulting values are close to those found in the literature [2.9]. For undoped, monocrystalline silicon, k_{Si} shows a $1/T^{1.3}$ temperature dependence. Section 2.5.2.4 will discuss the effect of doping in more detail.

2.2.3.3 Silicon Dioxide

From literature [2.12], the thermal conductivity of crystalline SiO_2 (also known as quartz) is similar to that of silicon. Since SiO_2 is an insulator, only phonons (and not electrons) contribute to its thermal conductivity. As for silicon, various phonon scattering mechanisms determine λ_{ph}, and thus k_{SiO2}.

However, due to the way in which they are fabricated, SiO_2 layers in IC technology are *amorphous* rather than *crystalline*. This greatly affects their thermal conductivity. In amorphous SiO_2, at above ≈100K, λ_{ph} reduces to the interatomic distance [2.1]: In the absence of a well-organized crystal lattice, each next atom is dislocated and thus can be considered a local defect [2.1], so that phonon-defect scattering strongly dominates. Because λ_{ph} is so small, k_{SiO2} is also small, which is in accordance with the well-known fact that glass (amorphous SiO_2) is a thermal insulator.

An interesting consequence of λ_{ph} being limited by the interatomic distance is that, beyond the Debye temperature of SiO_2, k_{SiO2} is nearly temperature-independent. This is confirmed by measurements in literature, e.g. in [2.11], in which the thermal conductivity of the SiO_2 between two metal layers in a CMOS process was found to only change from 1.1 to 1.2 W/m/K over the 0 to 125ºC range.

2.2.4 Thermal Diffusivity

Armed with a fundamental understanding of heat capacity and thermal conductivity, we can now investigate thermal diffusivity, D, the rate at which heat diffuses through a material. From the previous discussion on phonon transport, D can be considered as a 'thermal mobility' or a 'thermal time constant' that is obtained when the equation for k (2.6) is divided by the heat capacity. D (in m^2/s) is indeed the ratio of k and C_v and is given by:

$$D = \frac{k}{C_V} = \tfrac{1}{3} v_s \lambda_{ph} \tag{2.7}$$

The thermal diffusivity is thus only determined by v_s and λ_{ph}. D_{Si} and D_{SiO2} are straightforwardly obtained from the previous analyses and correspond well to data from reference measurements [2.12] Fig. 2.5 shows both historical data and simulated values for D_{Si} and D_{SiO2} based on the model of Holland [2.6]. D_{SiO2} is multiplied by 100x to fit on the same scale.

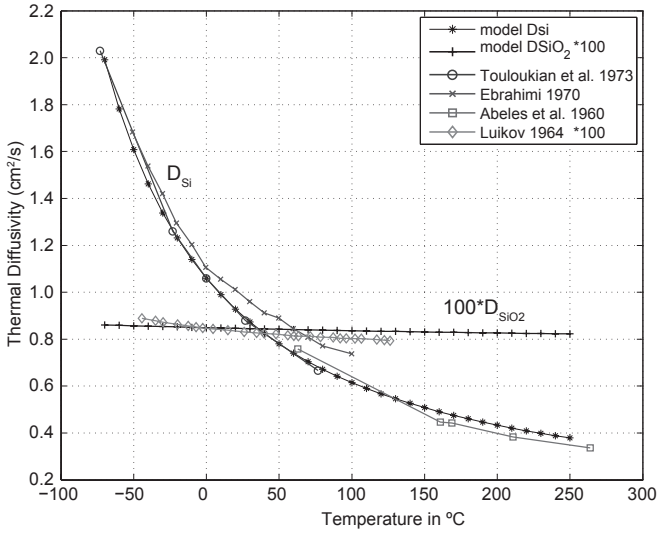


Fig. 2.5: Historical data and simulated values for D_{Si} and D_{SiO2}.

The first of two observations that can be made here is that D_{Si} is strongly temperature-dependent, while D_{SiO2} is much less so. This difference can be exploited to measure absolute temperature (as will be shown later in this thesis). Secondly, Eq. (2.7) shows that D is only technology-dependent in so far as IC processing affects v_s and λ_{ph}. The former being practically constant and λ_{ph} being only weakly sensitive to doping variations (see [2.7] and section 2.5.2.4), these thermal time constants can be expected to be very well-defined. This is the main motivation for the research described in this thesis.

In developing systems based on measuring D_{Si} and D_{SiO2}, it is useful to abstract the rather complicated underlying thermal models into approximate temperature dependencies. For D_{Si}, a $1/T^{1.8}$ power law proportionality holds well over the -55ºC to 125ºC range (D_{Si} at 27ºC = 0.8cm^2/s); D_{SiO2} can be assumed constant at D_{SiO2} = 0.008cm^2/s.

The electrical mobility μ can be considered analogous to the "thermal mobility", D. In fact, μ(T) can also be approximated by a $1/T^n$ power law, with n ≈ 1.1 – 1.3. In non-extrinsic semiconductors, the underlying physical principles are the same as those underlying the temperature dependence of k_{si}, i.e. μ is also limited by phonon scattering. However, μ is usually measured via the electrical conductivity (σ), which is ≈1000x more sensitive to process spread than k.

Although it is potentially useful that D is insensitive to process spread, the inverse is also true: the electrical characteristics of most IC components are only very weakly sensitive to changes in D. Measuring D, therefore, requires a different kind of component: an electrothermal filter.

2.3 Electrothermal Filters that measure D_{Si}

2.3.1 ETF Theory

Electrothermal filters, or ETFs, are components whose time-domain properties are determined by D. They consist of a heater, a thermal medium and a temperature sensor. Both their input and output are in the electrical domain, while the thermal-domain delay (or phase shift) between input and output is a measure of D.

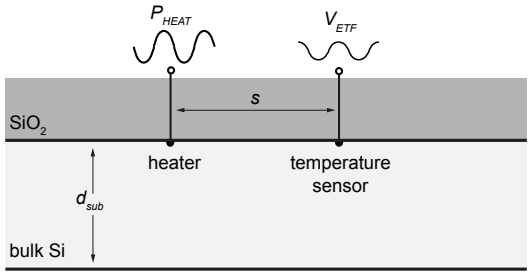


Fig. 2.6: A basic ETF that measures D_{Si}.

Fig. 2.6 shows the cross-section of an ETF that measures D_{Si}, i.e a 'silicon ETF'; it consists of a point heater and a point temperature sensor, realized on the surface of a silicon chip. For typical values of their spacing s < 50μm, the silicon substrate can be considered infinitely thick (d_{sub} ≈500μm). For now, the SiO_2 layer (formed by field and intermetal oxides) will be considered to be a perfect thermal insulator. When the heater is driven by a harmonic signal at an angular frequency ω, heat diffuses into the silicon semi-spherically. The associated temperature profile is given by solving [2.3]:

$$\frac{d^2(rT(r))}{dr^2} - q^2 rT(r) = 0\,, \tag{2.8}$$

In which r is the distance to the point source, $T(r)$ is the temperature increase at this distance and $q^2=j\omega/D$. The solution to Eq. (2.8) for a point heater [2.3] defines the thermal impedance, $Z_{th}(\omega,r)$, which relates $T(\omega,r)$, the temperature increase at the sensor, to $P_{heat}(\omega)$:

$$Z_{th}(\omega,r) = \frac{T(\omega,r)}{P_{heat}(\omega)} = \frac{1}{2\pi k_{Si} r}\exp(-r\sqrt{\frac{\omega}{2D_{Si}}})\exp(-jr\sqrt{\frac{\omega}{2D_{Si}}}) \tag{2.9}$$

$Z_{th}(\omega,r)$ has magnitude and phase components given by:

$$\begin{aligned} |Z_{th}(\omega,r)| &= \frac{1}{2\pi k_{Si} r}\exp(-r\sqrt{\frac{\omega}{2D_{Si}}}) \\ \phi(\omega,r) &= -r\sqrt{\frac{\omega}{2D_{Si}}} \end{aligned} \tag{2.10}$$

These equations show that $|Z_{th}|$ is a function of both k_{Si} and D_{Si} that decreases with distance. It also decreases for increasing ω, so ETFs are (thermal-domain) low-pass filters. When the temperature sensor has a sensitivity S_t, the ETF's output amplitude V_{ETF} is given by:

$$V_{ETF}(\omega,r) = \frac{P_{HEAT} S_t}{2\pi k_{Si} r}\exp(-r\sqrt{\frac{\omega}{2D_{Si}}}) \tag{2.11}$$

The phase shift of an ETF, ϕ_{ETF}, is defined as the phase shift at the temperature sensor when the heater is driven at a frequency f_{drive}. ϕ_{ETF} is given by:

$$\phi_{ETF} = -s\sqrt{\pi f_{drive} / D_{Si}} \propto -s\sqrt{\pi f_{drive} T^{1.8}} \tag{2.12}$$

Eq. (2.12) shows that ϕ_{ETF} is only dependent on geometry, driving frequency and D_{Si}. Note that, since phase is proportional to $\sqrt{f_{drive}}$, the thermal delay itself is frequency-dependent.

Eq. (2.10) shows that both V_{ETF} and ϕ_{ETF} are temperature-dependent, and thus both could potentially be used as the basis of a temperature sensor. V_{ETF} has several degrees of freedom, including P_{heat} and S_t, both of which are difficult to measure and/or control accurately. ϕ_{ETF}, on the other hand, is only dependent on f_{drive} and geometry (through s), and thus seems a better candidate for an accurate temperature sensor.

Since ϕ_{ETF} is the temperature variable of interest, the sensor that picks up thermal signal does not have to measure *absolute* temperature; rather, it only has to detect *relative* temperature fluctuations, simplifying its implementation. The sensor will typically drive a phase detector that extracts ϕ_{ETF} from its output. For noise and offset reasons, S_t should be maximized.

When the above ETF is driven at a constant f_{drive}, $\phi_{ETF}(T)$ is determined by D_{Si} and s. From section 2.2.4, ϕ_{ETF} will have a near-linear $T^{0.9}$ temperature dependence that is process-independent. Fig. 2.7 shows the simulated ϕ_{ETF} for s=24μm and f_{drive} = 85kHz in bulk CMOS[1]:

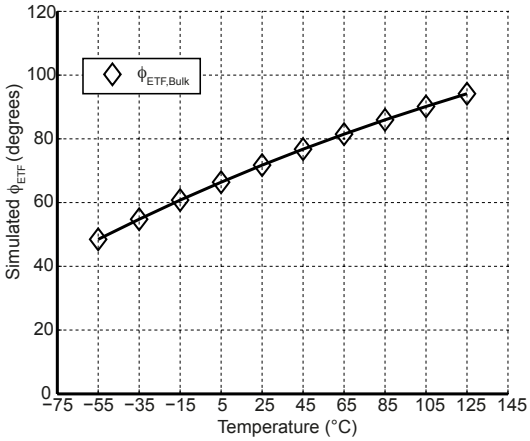


Fig. 2.7: Simulated ϕ_{ETF} as a function of temperature.

While the point-heater, point-sensor ETF offers useful basic insights into thermal diffusivity sensing, practical ETFs tend to be more complex. In practice, an important tradeoff is associated with the choice of s: given a certain absolute lithographic spread ds, s should be maximized for small ds/s and thus high accuracy, even though Eq. (2.11) shows that this reduces V_{ETF}. Quantifying and discussing this tradeoff in more detail requires a better understanding of how to implement ETFs in CMOS technology. As we will see, the need to minimize error and maximize resolution drives the design of the heater and the temperature sensor, and also determines their spacing.

2.3.2 Implementation

In CMOS technology, an ETF's heater and sensor can be implemented in several ways. However, for silicon ETFs, best performance can be expected when:

1. Heater and sensor are both directly implemented in the substrate, so that they are in good thermal contact with the heat diffusion path.

[1] Note that a thermal delay should, strictly speaking, correspond to a negative phase shift; throughout this thesis, ϕ_{ETF} is meant to indicate $|\phi_{ETF}|$

2. Their spacing is accurately defined; this is normally the case since geometries are generally defined by optical lithography, or by diffusions and their associated depletion regions. Lithography has the advantages of being insensitive to process spread (e.g. doping fluctuations) as well as benefiting from ever-increasing accuracy (through CMOS scaling). Using lithography to determine s thus seems favorable.
3. V_{ETF} is maximized for a fixed P_{HEAT} or, equivalently, P_{HEAT} is minimized for a fixed V_{ETF}. This implies that S_{tc} should be as high as possible, and that as much of the generated heat should contribute to V_{ETF}. Ideally, the heater should be completely surrounded by highly sensitive (relative) temperature sensors.
4. Most of P_{heat} should be concentrated in a single tone (at f_{drive}). In practice, driving the heater with a square wave is easier than using a sine wave, but the former's higher harmonics will be attenuated by the ETF's thermal low-pass filtering.

In CMOS technology, heaters can be implemented by any diffused component: resistors, diodes, MOSFETS and (parasitic) BJTs. Fig. 2.8 shows both resistor and MOSFET implementations.

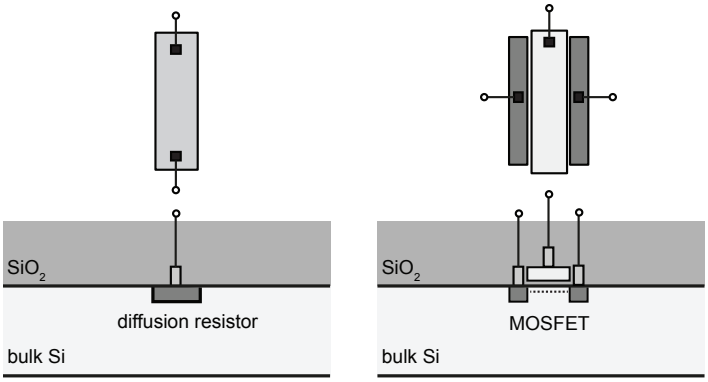


Fig. 2.8: Heater implementations using diffusion resistors or MOSFETS.

Resistors have a simple structure and generate heat (near-) uniformly, but they require driving circuitry to generate an AC P_{HEAT} signal. A MOSFET operating in triode also acts like a resistor, and is easier to drive (via its gate). However, its polysilicon gate upsets the previous assumption that the oxide can be considered a thermal insulator. Process spread in gate oxide layer and polysilicon layer characteristics may affect the thermal impedance seen by the heater, introducing error. Additionally, a MOSFET will have a composite heating profile consisting of channel and source/drain diffusion heaters, introducing a new degree of freedom and a possible source of process spread. The main focus of this work will be on using diffusion resistors as heaters, because of their simplicity.

Given a target P_{heat}, the geometry of a resistive heater is determined by its sheet resistance ($R_{\square}$) and the supply voltage, V_{DD}. Heaters should be as small as possible, since this produces the largest temperature excursions and thus the largest signals; a useful analogy is that of a tentpole raising a circus tent, and so even at some distance from the heater, temperature excursions are larger when the heater is small.

Having a point-heater is desirable from a theoretical point of view, but not realizable in practice. Typical values for P_{heat}, V_{DD} and $R_{\square}$ (for p^+- or n^+-diffusion resistors) require L>>W, in which case a point heater can be approximated by folding the heater resistor. In general, U-shaped heaters approximate a point heater sufficiently well. Some example heaters are shown below:

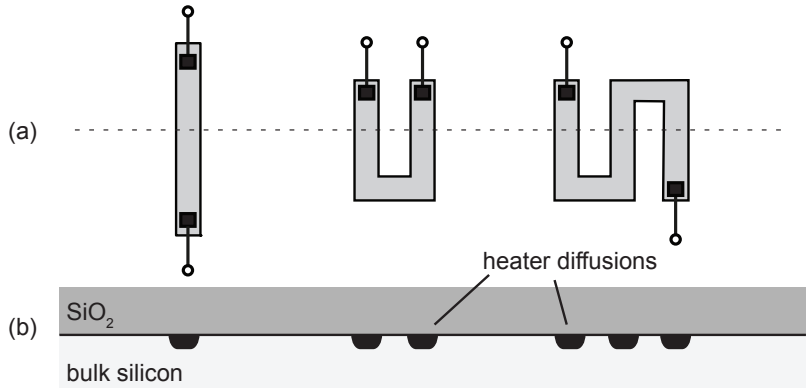


Fig. 2.9: Several diffusion resistor heater profiles (a) and their cross-section at the dashed line (b).

The relative temperature sensor can also be implemented by any diffused component, since these all exhibit temperature-sensitive characteristics. However, the effective location of diffusions is sensitive to process spread and will thus have an effective *s* that is also process-dependent. Moreover, diffusion resistors, MOSFETs or BJTs all need biasing, which leads to the potential drawbacks of extra power consumption and self-heating.

Another way to create a relative temperature sensor is by using the Seebeck effect, which states that a temperature difference between the two ends of two dissimilar conductors results in a voltage proportional to this temperature difference. The Seebeck effect can be used to create on-chip thermocouples, which can be series-connected to form thermopiles. Integrated thermopiles have been the subject of extensive research (see e.g. Van Herwaarden [2.13]). It has been found that the thermocouples that form between aluminum interconnect and diffusion resistors have a relatively high sensitivity, S_{tc}, of about 0.5mV/K for p^+-diffusion / Al thermocouples. This is much higher than for conventional metal/metal thermocouples, and this is due to the high Seebeck coefficient of the doped silicon [2.13]. These thermocouples form at the metallurgical junction and are thus defined by the contact plugs that connect the lowest metal layer to the diffusion surface. The geometry of these junctions is *only* defined by lithography and not by the consecutive diffusion implant. These thermocouples can thus be expected to have a well-defined geometry, making them a good candidate for use in ETFs. Also, thermopiles are passive sensors that do not require biasing, and they are intrinsically offset-free. For these reasons, this thesis will focus on thermopiles as the relative temperature sensor. Fig. 2.10 shows the layout of the resulting, more complex, ETF:

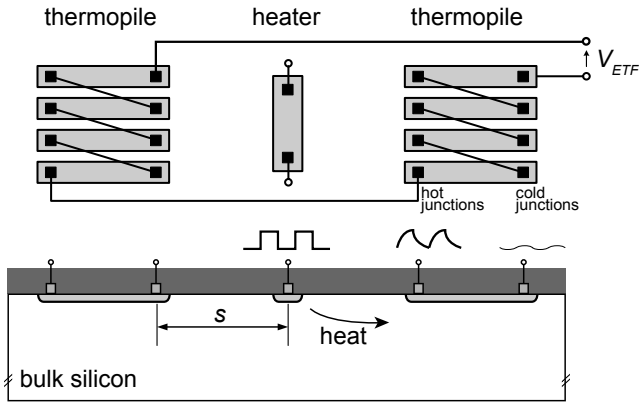


Fig. 2.10: A bar ETF using a thermopile as its relative temperature sensor.

In this ETF, the heater has been implemented as a simple resistive bar. The thermocouple junctions closest to the heater are the *hot* junctions, while the others are the *cold* junctions. The polarity of the detected signal at the cold junctions is opposite to that at the hot junctions, but due to the thermal filtering, its amplitude will be substantially reduced. V_{ETF} is given by:

$$V_{ETF} = n(\alpha_1 - \alpha_2)(T_{hot}(\omega) - T_{cold}(\omega)) \tag{2.13}$$

In this equation, n is the number of thermocouples, α_1 is the Seebeck coefficient of the diffusion resistor, and α_2 is the Seebeck coefficient of aluminum. T_{hot} and T_{cold} are the temperature signals at the respective thermocouples; these are obtained by modeling the CMOS ETF.

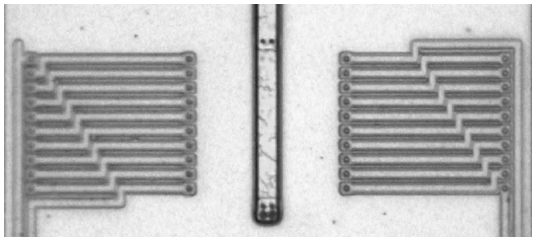


Fig. 2.11: Micrograph of a bar ETF [2.17].

2.3.3 Modeling

2.3.3.1 Numerical method

Practical ETF implementations are more complex than the point-heater, point-sensor ETF discussed previously, so deriving Z_{th} for such ETFs analytically is rather difficult, and in many cases impossible. Fortunately, a sufficiently accurate model can be developed through the use of finite-element techniques.

The thermal impedance of an ETF with an arbitrary heater shape can be determined by dividing the heater volume into a mesh of half-spheres having the same equivalent total volume, and summing the thermal impedance vectors from each of the thermocouple junctions to each of the half-spheres [2.14]. This is illustrated by Fig. 2.12:

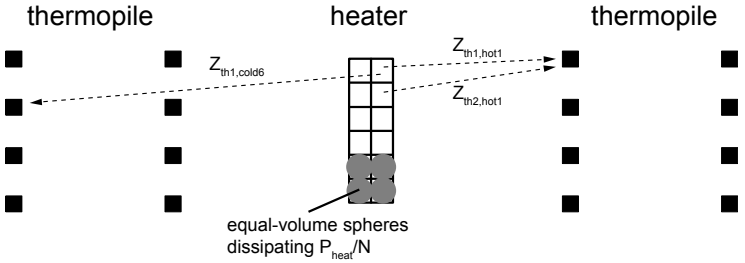


Fig. 2.12: Finite-element model for calculating an ETF's thermal impedance.

Outside of the heater half-spheres, Eq. (2.11) holds, in which case the total thermal impedance is given by:

$$Z_{th,ETF} = \frac{1}{N}\sum_{n=1}^{N}\sum_{p=1}^{P} Z_{th,hot}(r_{n,p}) - \frac{1}{N}\sum_{n=1}^{N}\sum_{p=1}^{P} Z_{th,cold}(r_{n,p}) \tag{2.14}$$

In this equation, N is the number of heater segments, P is the number of hot (and cold) junctions, and $r_{n,p}$ are the distances between each of these elements. The total thermal impedance, $Z_{th,ETF}$, is the sum of all these thermal impedances.

2.3.3.2 Phase-contour ETFs

$|Z_{th}|$ is maximized when all hot junctions see the same signal, which is the case when they are aligned on an equal-phase contour [2.15], or, equivalently, when all hot junctions each see the same thermal impedance. Fig. 2.13 shows the top view of a typical CMOS ETF with s=24μm and 24 thermocouples, 12 on each side of the U-shaped heater.

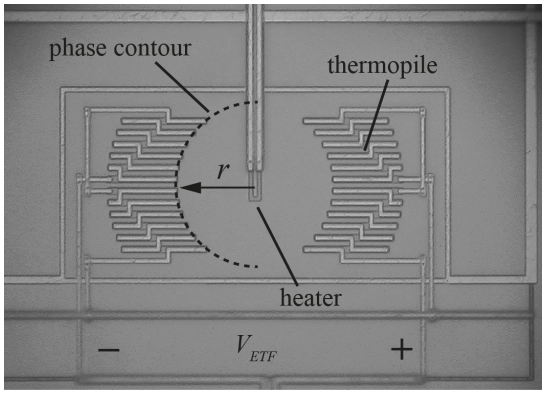


Fig. 2.13: Silicon ETF with thermocouple junctions aligned on a phase contour.

Advantageously, phase-contour ETFs have maximal $|Z_{th}|$ and thus output the largest signal for a fixed P_{heat} and s. Moreover, their ϕ_{ETF} can be approximated using the same equation used for the point ETF (Eq. (2.10)). This results in ϕ_{ETF} being proportional to $sf^{0.5}T^{0.9}$.

Summarizing this section, there are several ways to implement ETFs, and phase-contour ETFs and bar ETFs using resistive heaters and thermocouple sensors appear to be promising basic sensor structures. Their key properties (e.g. accuracy, SNR, scaling behavior) are discussed in more detail in section 2.5. This thesis first presents a similar overview for oxide ETFs.

2.4 ETFs that measure D_{SiO2}

2.4.1 Technology

Analogous to silicon ETFs, in which heat diffuses through the silicon substrate, heat generated in an ETF that measures D_{SiO2} should diffuse through a silicon dioxide layer. In bulk CMOS technology, several oxides can be used to measure D_{SiO2}:

1. gate oxide: the thin oxide layer separating the polysilicon gate from a MOSFET channel
2. field oxide / shallow trench oxide: the SiO_2 that separates polysilicon resistors from the substrate and separates surface junctions.
3. intermetal oxides: SiO_2 deposited between different metal layers

And, in the case of SOI technology:

4. buried oxide: the layer that isolates the epitaxial silicon layer from the substrate
5. deep trench isolation: the structures that electrically isolate each silicon ‘island’ / ‘tub’.

Several of the above oxides cannot be used because the heater and temperature sensor of an ETF have to straddle the thermal medium. This excludes the use of the buried oxide layer in SOI, since no electrical components can be created below it. Sensing through the intermetal oxide is also difficult,

because the metal layers are usually so low-ohmic that it is difficult to use them as heaters or as thermistors. In some processes, however, thin-film resistors could be used.

The gate oxide is also not an attractive option, because it is very thin. As a result, the associated thermal delay will be small (tens of nanoseconds) and so will be hard to measure, and will be affected by the thermal properties of the polysilicon gate and bulk silicon. Moreover, in advanced processes, the gate may be engineered for electrical performance, which will also affect its thermal properties, resulting in spread.

Consequentially, the most promising candidates for measuring D_{SiO2} are the field oxide and, in SOI technology, the oxide found in deep trench isolations.

Field oxide is an amorphous SiO_2 layer with a thickness of about 300nm (although this is technology-dependent). In older technologies, it is thermally grown through the local oxidation of silicon (the LOCOS method). More advanced CMOS technologies use shallow trench isolation (STI), in which trenches etched in the wafer surface are filled by depositing an amorphous SiO_2 layer on the wafer. Both of these oxides are shown in Fig. 2.14:

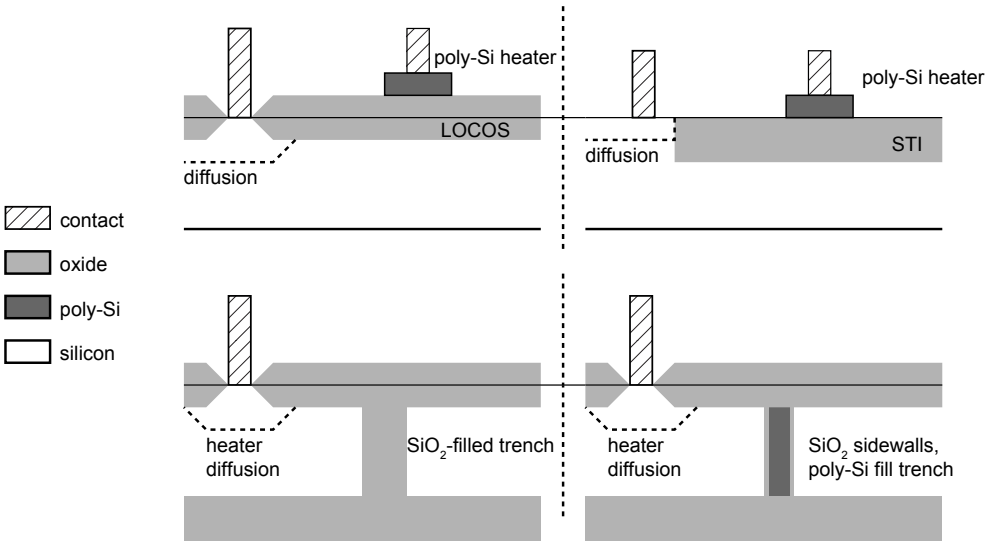


Fig. 2.14: Various SiO_2 layers as an ETF's thermal medium.

Deep trench isolation structures are generally made in two ways: either the trench is completely filled with SiO_2, or the trench only has SiO_2 sidewalls and is filled with polysilicon; both structures are schematically drawn in Fig. 2.14. The second variant, a 'sandwich' structure, is also shown in Fig. 2.16. The thermal delay of such a trench is still mostly determined by the oxide sidewalls, since the thermal diffusivity of polysilicon is comparable to that of silicon (it is about 2x lower due to increased phonon-boundary scattering on the polysilicon grain boundaries [2.10]).

2.4.2 Implementation

When making D_{SiO2} ETFs, it is difficult to create a thermal path that is not affected by the large volume of highly thermally conductive bulk silicon (without resorting to methods like etching cavities).

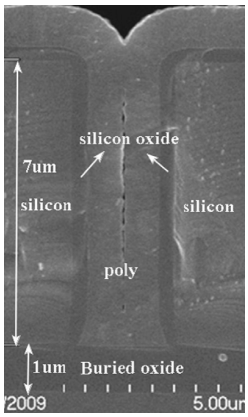


Fig. 2.16: Typical SiO_2-sidewall trench, after [2.30], with permission.

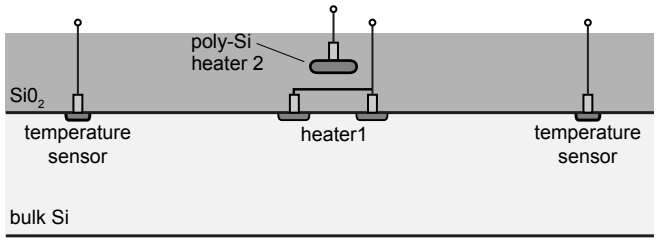


Fig. 2.15: An ETF with multiple heaters to measure ϕ_{ox}.

This is especially true since most heaters and/or temperature sensors are made in the substrate, where they are in good thermal contact with the silicon. In order to determine D_{SiO2} then, an indirect approach using two ETFs can be used.

Fig. 2.15 shows a cross section of an ETF with two heaters; one implemented as a polysilicon resistor, the other as a diffusion resistor. Heat generated in the polysilicon heater will first have to diffuse through the field oxide, before spreading radially through the bulk silicon as in a silicon ETF.

Denoting the ETF's phase shift by ϕ_{ETF2} when heater 2 is driven, and by ϕ_{ETF1} when heater 1 is driven, then $\phi_{ETF2} - \phi_{ETF1} = \phi_{ox}$, the phase shift associated with heat diffusion through the field oxide.

For a ϕ_{ox} measurement based on deep trench isolation, a similar approach can be used. The phase shift difference between the two ETFs shown in Fig. 2.17 will be determined by the thermal delay of the trench, i.e. $\phi_{ETF2} - \phi_{ETF1} = \phi_{trench}$:

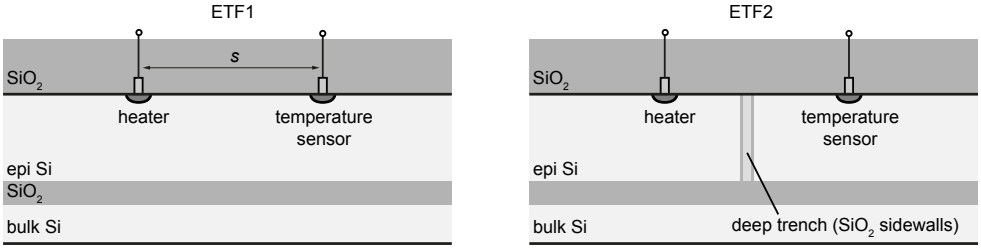


Fig. 2.17: Two ETFs used to measure ϕ_{trench}.

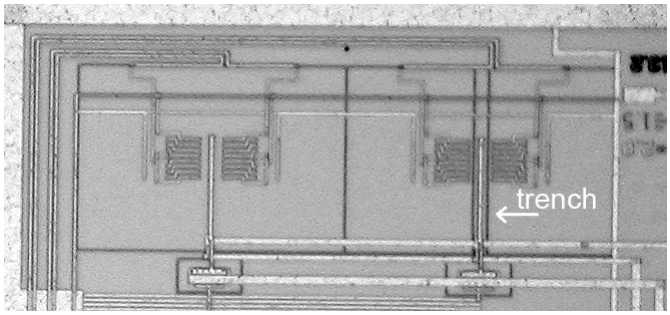


Fig. 2.18: Micrograph of two-ETF ϕ_{trench} sensor.

Fig. 2.18 shows a micrograph of a two-ETF ϕ_{trench} implementation; the dark vertical lines are the deep trench isolation structures. The bar-ETF approach was used for simplicity and compliance to the manufacturing rules of the process used.

2.4.3 Modeling

Modeling oxide ETFs is non-trivial, because closed-form analytical expressions for heat conduction become very complex for ETFs consisting of multiple materials with different geometries. Although finite element methods could be used to develop insight, this is computationally intensive: ϕ_{ETF} needs to be calculated at many ambient temperatures, at many values of f_{drive}, for ETFs having a geometrical resolution of about 0.01%. This was not done in this research, but is an interesting subject for future work.

A more readily useable model for oxide ETFs is based on empirical techniques. Consider the ETF of Fig. 2.15, in which the heater is implemented by a polysilicon resistor and thus sits on top of the field oxide, while the temperature sensor is still spaced at distance s. Compared to the D_{Si} ETF discussed before, heat now first has to diffuse through the field oxide over a distance t_{ox}, before diffusing radially into the substrate. This structure can be considered as two series-connected ETFs, having a total thermal path length equal to $s + t_{ox}$ and a phase shift given by.

$$\phi_{ETF2} = \sqrt{\pi f_{drive}} \left(\frac{s}{\sqrt{D_{Si}}} + \frac{t_{ox}}{\sqrt{D_{SiO2}}} \right) \tag{2.15}$$

This model holds reasonably well for $s >> t_{ox}$, which is the case for most ETFs and can be guaranteed by design. ϕ_{ox} can now be determined indirectly, by evaluating the difference between an ETF with a polysilicon heater and a diffusion heater:

$$\phi_{ox} = \phi_{ETF2} - \phi_{ETF1} = t_{ox} \sqrt{\pi \frac{f_{drive}}{D_{SiO2}}} \tag{2.16}$$

This equation shows that, compared to silicon ETFs, the phase shift of oxide ETFs is similarly dependent on f_{drive} and D. Fig. 2.19 shows simulation results for two s = 8µm ETFs, with one of them having a trench between heater and sensor with a total sidewall thickness of 300nm, and thus having a larger ϕ_{ETF}. Following Eq. (2.16), the difference in ϕ_{ETF} is defined as ϕ_{ox}. As can be seen, ϕ_{ox} has a nominal value that is comparable to ϕ_{Si}, but substantially less temperature dependent. This is in line with expectations based on D_{Si} and D_{SiO2} and the geometry of the structures.

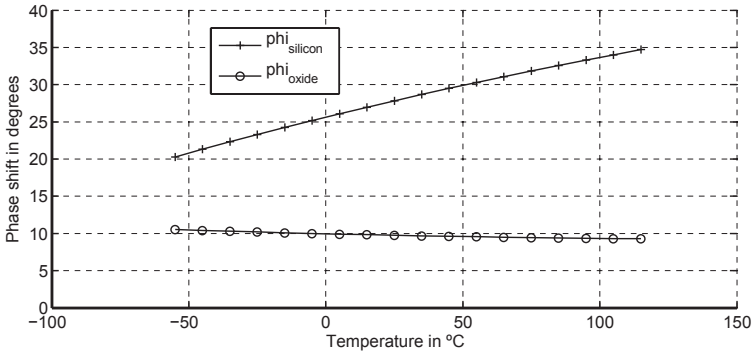


Fig. 2.19: Simulated phase shift for silicon and silicon dioxide ETFs over temperature, based on the modeled values for D_{Si} and D_{SiO2} (see e.g. Fig. 2.5).

2.5 Aspects of ETF Performance

2.5.1 Overview

Having developed a general understanding of the characteristics of silicon and oxide ETFs in the preceding sections, the focus will now be shifted to several of their properties that are important to diffusivity-based temperature sensing.

Although ETFs and other integrated thermal-domain components have been studied in the past (see section 1.4.3), their application to temperature sensing is relatively new. The existing body of work consists mainly of the work of Szekely [2.21][2.22], who showed that D_{Si}, when sensed by a thermal oscillator, is indeed temperature-dependent. The works of Witte, Snoeij and Makinwa later show significant improvements to inaccuracy, power consumption and interfacing circuitry[2.25][2.26].

While the prior art thus shows that ETFs have potential as temperature sensors, they should be studied further and in greater detail in order to answer the main research question of this thesis, i.e. whether ETFs can be *competitive* temperature sensors.

As outlined in Chapter 1, a competitive temperature sensor should have a high accuracy, a high resolution, a wide operating range and a small size. These aspects will now be discussed individually, further developing the theoretical framework of sections 2.3 and 2.4 as needed. The ETF's scalability and compatibility with modern CMOS technology will be included in the discussion where relevant.

The main focus of this discussion is on silicon ETFs and their potential for scaling (i.e. improved performance in advanced IC technology), but where applicable, the oxide ETF will also be discussed. The following discussion will assume the availability of precision sensor interfaces. These will be discussed in greater detail in chapters 3 and 4.

2.5.2 Accuracy

2.5.2.1 Do the measured values for D_{Si} and D_{SiO2} correspond with literature?

Although the values of D and k for bulk silicon and silicon dioxide can be derived from solid-state physics, there is some disagreement in the existing literature about their experimentally determined values. At room temperature, measured values for D_{Si} ranging from 0.6cm^2/s [2.14] to 0.9cm^2/s [2.12] have been reported. The wide spread of these values may be due to variations in the samples under study, or the limited accuracy of traditional measurement methods (such as IR interferometry or the 3ω method) to establish D_{Si}.

ETFs offer the possibility of accurately determining the effective D of silicon and silicon dioxide structures. For a silicon ETF, for example, the measured ϕ_{ETF} can be input to an 'inverted' analytical ETF model [2.16]. Since f_{drive} is accurately known (to within a few ppm) and the average s can be assumed to be very close to the nominal design value (or measured using e.g. an electron microscope), D can be accurately determined from measurements. Note that this is an *effective* diffusivity, which will not be exactly the same as the bulk value. For instance in the case of a silicon ETF, it has been shown that the SiO_2 layer at the silicon surface also conducts a small part of the heat, which leads to a higher effective D [2.19]. By controlling P_{heat} and accurately measuring V_{ETF}, k can also be determined.

Determining the effective values for D and k (and their temperature dependence) is useful for e.g. better estimates of both AC and DC self-heating of integrated circuits. Especially in SOI technology, a good understanding of dynamic self-heating can be critical to performance and reliability.

In chapter 5, effective values for D_{Si} and D_{SiO2} over temperature will be presented.

2.5.2.2 What is the effect of lithographic misalignment?

Sensitivity to lithographic error

Under the assumption that D_{Si} is insensitive to process spread and f_{drive} is constant and known, the dominant source of error in ϕ_{ETF} (and thus the dominant temperature error) will be that of variations in effective s, due to lithographic misalignment.

Partial derivatives of Eq. (2.12) show that a constant ds leads to a slightly temperature-dependent dT:

$$\frac{\delta\phi_{ETF}}{\phi_{ETF}} = \frac{\delta s}{s} = -0.5\frac{\delta D}{D} = 0.9\frac{\delta T}{T} \tag{2.17}$$

For a silicon ETF with s=20µm, a ±0.5ºC error at 125ºC corresponds to a ds of ±23nm. While small, such tolerances are well within the reach of modern CMOS technology and continue to improve as CMOS scales down. This is one of the main reasons why silicon ETFs are attractive for thermal management in deep submicron CMOS chips.

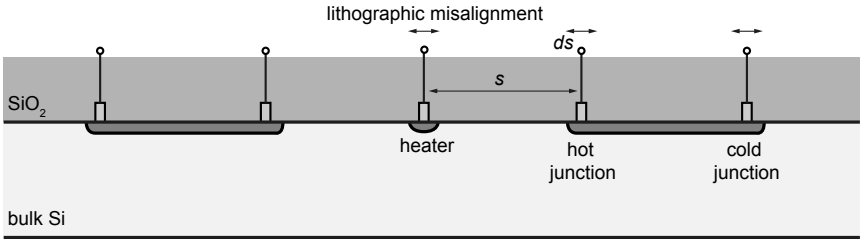


Fig. 2.20: ETF side view indicating sources of lithographic misalignment.

Sources of lithographic error

Typical ETFs have three sources of lithographic misalignment:

- Misalignment of the heater
- Misalignment of the hot junctions
- Misalignment of the cold junctions

These error sources have different physical origins. When the heater is a diffusion resistor, its center of mass is determined by the contact mask, the diffusion implant and, to lesser extent, the implant around the diffusion (namely through depletion layer modulation). Of these three degrees of freedom, especially implant dose and location is not well-controlled, so the heater will not have an accurate geometry. The use of (circular) symmetry, i.e. surrounding the heater with thermocouples, advantageously suppresses the effects of heater misalignment.

Misalignment of the cold junctions is a more substantial source of error, since this affects the effective s (s_{eff}). However, because the thermal signal at those junctions has already been substantially filtered by the die, its contribution is still relatively small.

At the hot junctions, however, this signal is largest, and there a position error corresponds to a linear error in s. Symmetrical layout reduces the effect of mask shifts (which is then effectively heater misalignment), but a residual random error remains. Assuming a normal distribution, a quadratic increase in the number of thermocouples results in a linear reduction in spread , but, in practice, the total number of thermocouples is bounded by layout constraints and parasitics.

It is of interest to understand which masks are most important in defining s accurately. In principle, the thermocouple junctions are created at the metallurgical junction between the diffusion and the low-ohmic contact, i.e. right at the surface of the die. In older CMOS technology, the contact directly

connects to the diffusion and the contact mask is the dominant source of lithographic spread. Newer CMOS technology typically uses a low-ohmic silicide layer between contact and diffusion, (driven by the need to reduce MOSFET terminal resistance). This layer (Cobalt silicide, CoSi, is common) is near-metallic, so that the difference in Seebeck potential between aluminum and silicide can be neglected. In silicided processes, therefore, the hot junctions are defined by the silicide mask. In both cases, the contact locations are defined by a single mask. Fig. 2.21 shows a comparison:

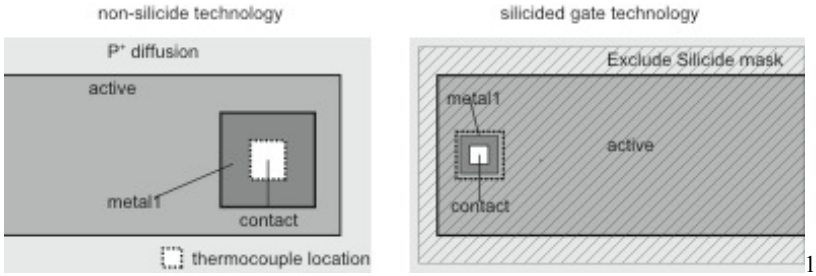


Fig. 2.21: Contact definition in silicided and non-silicided CMOS technology.

Simulation results

The simulations shown below are based on the ETF in Fig. 2.13, i.e. a phase-contour ETF with s=24µm having 24 hot junctions driven at f_{drive}=85kHz. The cold junctions are at a phase-contour with s=45µm. All sources of spread are assumed to have a Gaussian distribution and zero mean.

Fig. 2.22 shows the simulated error in ϕ_{ETF} for random spread in (x,y) heater coordinates with 6σ = 1µm (100 runs). This excessive amount of spread was chosen on purpose, since the effect of heater misalignment is expected to be small. Since any submicron CMOS process can be expected to have significantly better lithographic accuracy, this is not a significant source of spread.

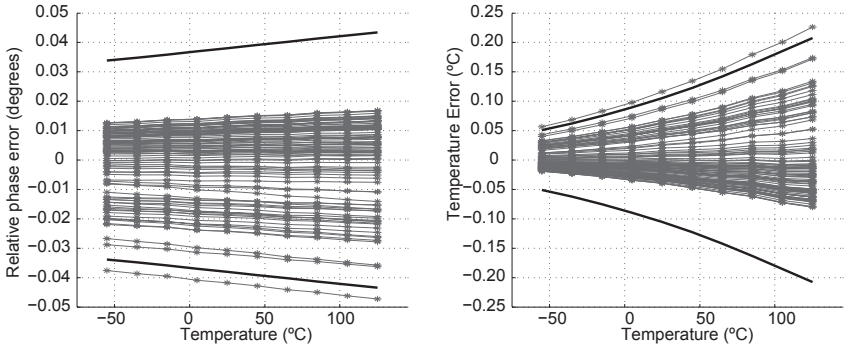


Fig. 2.22: Simulated phase and temperature errors due to random heater location variation; the thick lines indicate ±3σ limits.

Fig. 2.23 shows a simulation with random spread in the cold junction locations with 6σ = 0.18µm (100 runs). Interestingly, the relative contribution of cold junction error increases at low temperatures, because increasing D_{Si} and k_{Si} nonlinearly reduces the low-pass filtering between hot and cold junctions versus between the heater and the hot junctions.

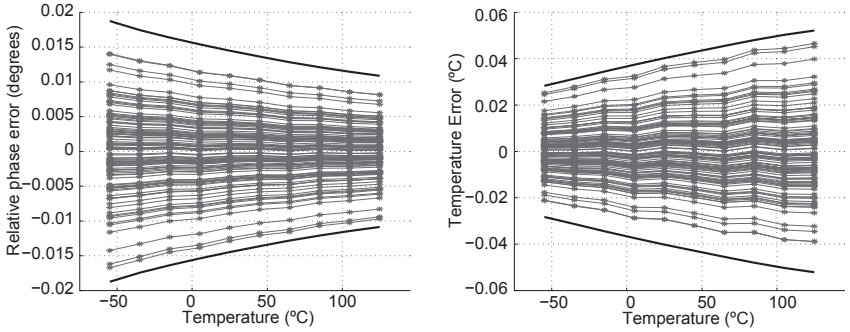


Fig. 2.23: Simulated phase and temperature errors due to cold junction location variation; the thick lines indicate ±3σ limits.

And finally, Fig. 2.24 shows the simulated error for hot junction misalignment ($6\sigma = 0.18\mu m$):

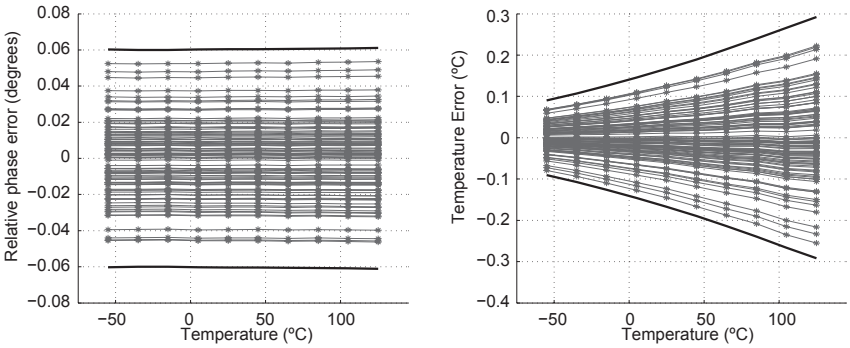


Fig. 2.24: Simulated phase and temperature errors due to hot junction location variation; the thick lines indicate ±3σ limits.

As expected, misalignment in the hot junctions is the dominant source of error and causes a constant relative phase error, which translates to a temperature-dependent temperature error.

2.5.2.3 Scaling

If lithographic spread is the dominant source of temperature error in ETFs, then ETF inaccuracy should decrease in more advanced, deep submicron CMOS processes. A *constant-size* ETF has constant s and should thus have a temperature error spread that is linearly proportional to the minimum feature length of the process, λ_{min}. Alternatively, an ETF with $s=n\,\lambda_{min}$ (with n a constant) should have a near-constant error spread, while reducing in area near-quadratically. Such *constant-error* ETFs do not get more accurate in more advanced technology, but section 2.5.3 will show that smaller s leads to significantly better SNR.

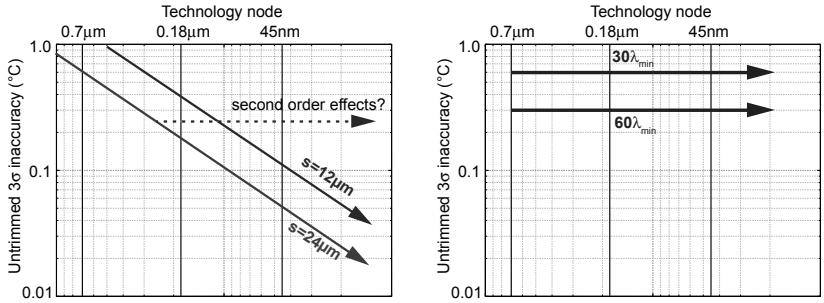


Fig. 2.25: Temperature-sensing inaccuracy as a function of technology node and spacing *s*, for constant-size and constant-error ETFs.

Fig. 2.25 shows the two approaches to ETF scaling in advanced CMOS technology. Note that only constant-size ETFs will potentially exhibit second-order effects that limit further improvement in temperature-sensing inaccuracy. Effects such as spread in D_{Si} or errors introduced by the readout circuitry will set a lower bound to inaccuracy. In constant-error ETFs, spread in s should always dominate these errors. The measurements that will be presented in chapter 5 will show results for both constant-size and constant-error ETFs.

2.5.2.4 Doping sensitivity

As discussed in section 2.2, D_{Si} is only weakly sensitive to fluctuations in doping. Only through the influence of phonon-impurity scattering does λ_{ph} decrease, and this is most notable at temperatures far below the military temperature range.

However, at high doping concentrations, phonon-impurity scattering does become a significant source of error. The effect of dopant atoms on D has been analyzed in Asheghi et al [2.7], in which a model for the doping sensitivity of k_{Si} is provided. This model has been used to calculate the effective value of λ_{ph}. Fig. 2.26 shows the resulting plots for D and ϕ_{ETF}:

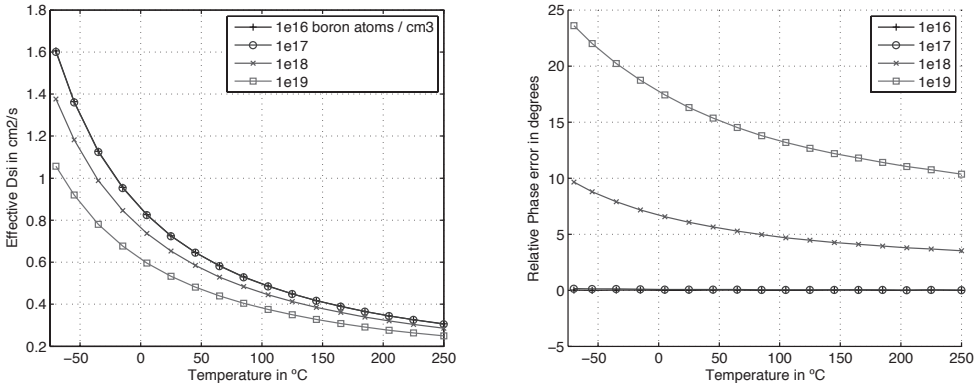


Fig. 2.26: Simulated D_{Si} and relative ϕ_{ETF} error as a function of boron doping.

While Fig. 2.26 shows that large doping levels do indeed cause a significant error in the nominal value of ϕ_{ETF}, this shift could potentially be dealt with through batch calibration. Random device-to-device error would require individual calibration, which is more elaborate. The effect of random doping fluctuations was simulated and the 3σ phase and temperature errors are given below (100 runs at each doping level with 3σ=20%);

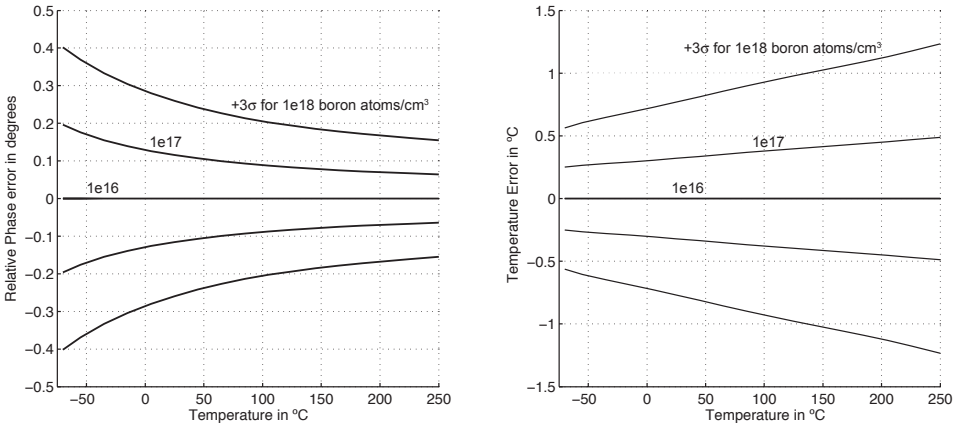


Fig. 2.27: Simulated effect of doping fluctuations around a nominal level. The lines indicate ±3σ limits.

This figure shows that ETFs in which the thermal path is lowly doped (e.g an epi layer) will indeed not have significant doping sensitivity. For the 0.7µm CMOS process used as one of the test substrates in this work, the doping level of the p-substrate is about $N_a = 5*10^{14}/cm^3$, while the doping level of the nwell is about $N_D=3*10^{16}/cm^3$. The above suggests that ETFs in such a technology will not be strongly sensitive to doping fluctuations. However, highly doped substrates or diffusions might cause a significant residual error in ϕ_{ETF}.

While the above model provides a useful first-order estimate, the actual sensitivity to doping is best tested by measuring ETFs with more strongly doped thermal paths. Fig. 2.28 shows two such ETFs; in one, heat diffuses through an n-well, while the other uses an n^{++}-implant to increase the doping level of the thermal path. Chapter 5 will present and discuss measurements results.

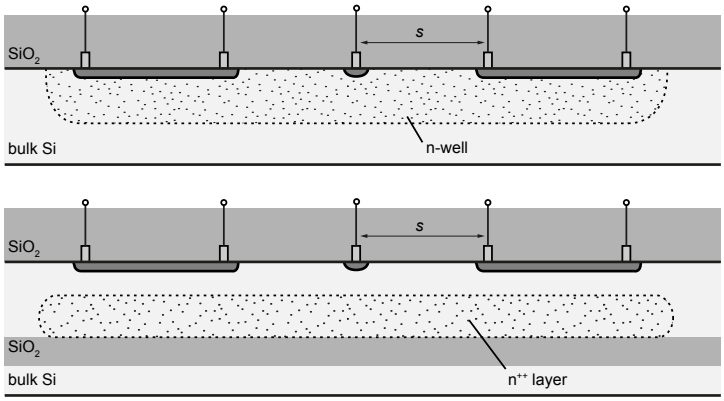


Fig. 2.28: Test ETFs to study the effect of doping on ϕ_{ETF}.

2.5.2.5 Wafer-to-wafer and batch-to-batch spread

Due to tolerances in the IC fabrication process, integrated electrical components exhibit spread between different process lots (also known as batch to batch spread). Batch-to-batch variation affects e.g. doping levels (and therefore resistivity) and gate oxide thickness; because of this variability, the absolute accuracy of IC components is limited to several tens of percent.

The effect of lot-to-lot variations on ETF is unknown. Section 2.3.1 has suggested that the dominant source of error is that of random lithographic variations in *s* that have zero mean; here, a batch-to-batch component is unlikely.

There are several second-order mechanisms through which ETFs might still be sensitive to process variations; three are discussed below:

1. It has been hypothesized that for IC substrates in which a lowly doped (p^-) epi-layer is grown on top of a highly doped (p^{++}) substrate, spread in the thermal conductivity of the substrate due to spread in doping may significantly affect the effective *D* [2.19].
2. The parallel heat diffusion path through the field and intermetal oxide having process-dependent thickness might affect the effective *D*.
3. In CMOS technology using STI, diffusions are typically more shallow than the field oxide; variations in STI thickness might modulate the effective *s* and thus cause a gain error. However, there is little vertical heat flow through the STI opening, so this error may not be very pronounced.

In conclusion, it is expected that the errors in ETFs are largely due to lithographic spread, but the above affects will be studied by evaluating the characteristics of ETFs realized in several wafers and different process runs.

2.5.2.6 Self-heating

Since ETFs need to produce heat to measure temperature and it is fundamentally impossible to produce negative heat in a resistor or a MOSFET, all ETFs self-heat when operating. This has two consequences:

- The DC component of P_{heat} will cause a DC temperature gradient across the thermopile. This causes an offset in V_{ETF}, which will have to be processed by the readout circuit.
- The effective D_{Si} is determined by the temperature profile along the heat diffusion path. As *s* is reduced to improve SNR, D_{Si} becomes more sensitive to spread in self-heating, thus introducing a new source of error.

The DC offset across the thermopile is straightforwardly analyzed by solving the heat equation for a DC input. For heat diffusing from a point heater to a point sensor in a semi-sphere of silicon, the DC temperature increase at radius *r* is given by:

$$T_{DC}(r) = P_{heat} * Z_{th}(r) = P_{heat} \frac{1}{2\pi k_{si} r} \tag{2.18}$$

Ignoring the temperature fluctuations at the cold-junction, the temperature difference across the thermopile is shown in Fig. 2.29:

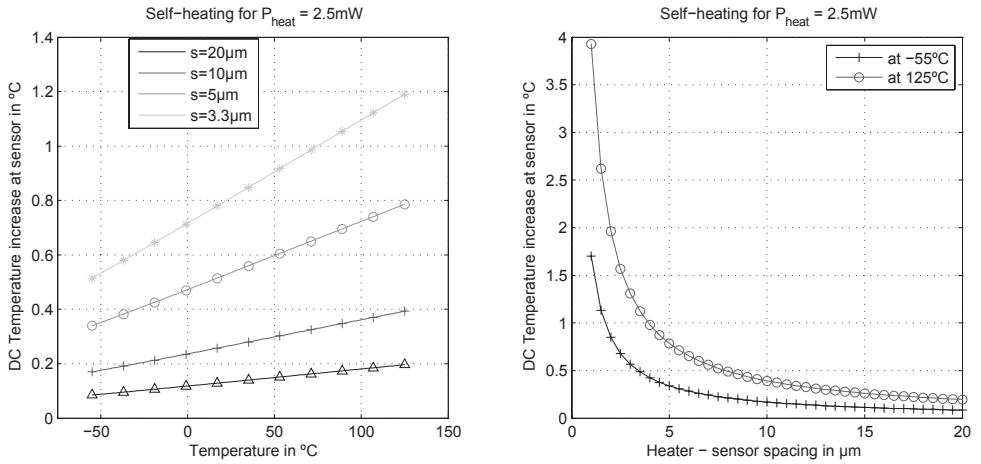


Fig. 2.29: Temperature increase at the hot junction as a function of T and *r*.

Assuming a thermopile with 24 junctions with an S_{tc} of 0.5mV/K, the offset voltage at the thermopile output will be up to 2.5mV for an s=20µm ETF. This will increase for scaled ETFs and might require offset-cancellation techniques in the readout circuitry.

The effect of self-heating on D_{Si} is less straightforward. The effective D_{Si} for a point-heater, point-sensor ETF is given by:

$$D_{si,eff} = \frac{1}{r}\int_0^r D_{si}(r)\,dr \tag{2.19}$$

$D_{Si}(r)$ can be approximated by the average temperature profile along the heat diffusion path, which leads to:

$$D_{si,eff} = \frac{1}{r}\int_0^r D_{Si}(T_{die} + \frac{P_{heat}}{2\pi k_{Si} r})\,dr \tag{2.20}$$

This can be numerically evaluated. Fig. 2.30 shows the resulting temperature error as a function of temperature for four different ETF spacings, i.e. 20, 12, 5 and 3.3µm, at a nominal P_{heat} of 2.5mW with ±20% variation. Interestingly, while the DC temperature difference across the thermopile is strongly temperature-dependent (as shown by the right-hand graph of Fig. 2.29), the temperature error in ϕ_{ETF} is near-constant over temperature.

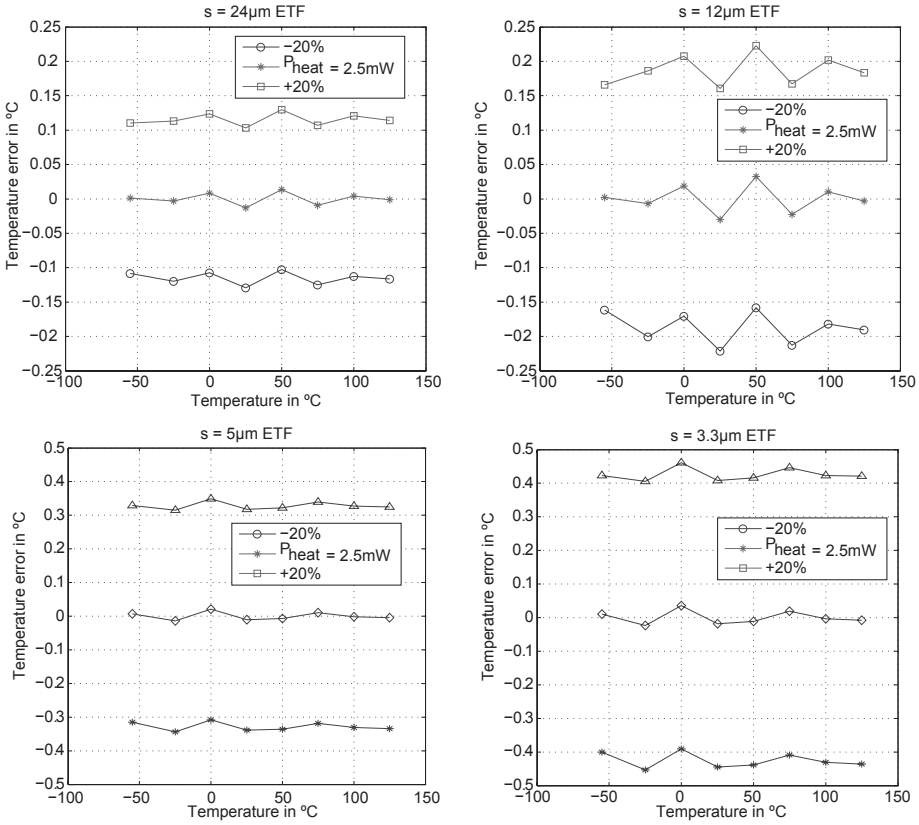


Fig. 2.30: Simulated temperature error due to spread in P_{heat}, as a function of temperature.

Self-heating is the main way in which the PSRR of ETF-based temperature sensors may be limited. In typical applications, the heater will be driven by a digital square wave with a peak-to-peak swing equal to V_{DD}. Variations in V_{DD} or heater resistance will change the effective D_{Si} around the heater, thus introducing temperature error. This will be more pronounced in SOI technology, since $k_{Si,SOI}$ is much smaller than in bulk CMOS. Fundamentally, ETFs with a better SNR will have a worse PSRR.

To improve PSRR, P_{heat} can be generated by a constant-power circuit that keeps V_{heater} * I_{heat} constant. Another solution would be to measure the difference in ϕ_{ETF} between two ETFs with different s, which will be much less V_{DD}-dependent. This is also the case for D_{SiO2} measurements based on two ETFs: ϕ_{ox} is only affected by the *difference* in self-heating. This should greatly improve their PSRR.

As will be shown in the measurements section, self-heating can be studied by sweeping P_{heat} and calculating the resulting ΔT and associated PSRR.

2.5.2.7 Thermal interference

When ETFs are used to measure the die temperature of large mixed-signal chips, AC heat generated by other circuitry can cause thermal interference. The power cycling of large digital circuit blocks or switching power devices will generate AC thermal signals in the substrate. The phase, frequency and amplitude characteristics of these signals may be unknown and potentially cause errors in ϕ_{ETF}, increasing inaccuracy. Fig. 2.31 shows an ETF next to circuitry that produces interfering heat:

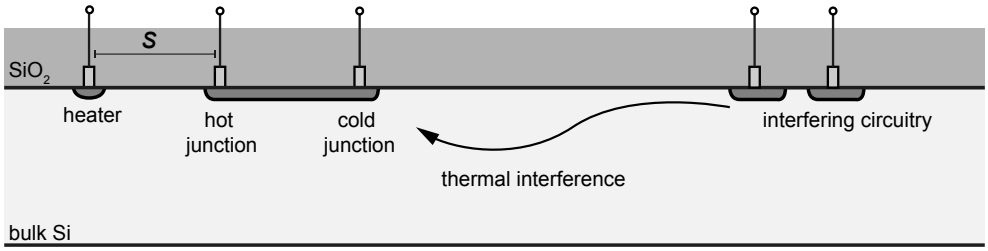


Fig. 2.31: Side view of an ETF in the presence of thermal interference.

There is a fundamental difference between BJT-based and ETF-based temperature sensors which is relevant in case of thermal interference. BJTs are typically read out by an ADC that samples and averages their output signal, while ETFs fundamentally require demodulation to extract ϕ_{ETF} from V_{ETF} and only then perform averaging to reduce thermal noise.

Because of this mixing operation, interference at f_{drive} or its odd multiples will cause errors in the extracted value of ϕ_{ETF}. If these signals have a constant phase relation to f_{drive}, averaging the demodulated output will not help to filter out these errors. Note that thermal signals at frequencies (much) lower than f_{drive} are actually the signals of interest (corresponding to self-heating transients); these are detected in much the same way as for BJTs.

There are a several ways to mitigate the effect of interference. First, increasing the distance between the ETF and the thermal aggressor reduces interference, since the silicon substrate will low-pass filter thermal signals with an approximately -30dB/decade slope (proportional to $r^{1.5}$). However, in dense digital circuitry, it might not be possible to locate ETFs away from digital gates, and defining a keep out area would greatly increase the effective area of ETFs, making them less attractive. Secondly, scaling the ETF by reducing s and increasing f_{drive} also reduces sensitivity to interference: ϕ_{ETF} is based on local, AC temperature fluctuations, and reducing s makes fluctuations more local. Coincidentally, this means that scaled ETFs such as those envisioned for thermal management in deep submicron CMOS technologies will not only have better SNR, they will also be less sensitive to interference.

If both spacing and scaling are not sufficiently effective, other solutions could be as follows.

1. Increasing P_{heat} to overcome the interference.
2. Increasing f_{drive} to increase the low-pass filtering action of the substrate on external agressors at f_{drive}.
3. Driving the heater and the V_{ETF} demodulator with a spread-spectrum clock, so that the effect of a single frequency disturbance is reduced.
4. Adding a watchdog circuit to detect unusual amplitude variations in V_{ETF} and thus detect erroneous signals.

One way to evaluate the effect of thermal interference is to drive a parasitic heater near the ETF with an interfering signal and measure the effect on ϕ_{ETF}. The measurements section will present the results of such an experiment.

2.5.2.8 Mechanical Stress

Encapsulating chips in plastic packages causes mechanical stress gradients across the silicon die and local stress (e.g. due to filler granules). This affects BJT-based temperature sensors, because their saturation current (I_s) is stress-dependent due to the piezojunction effect [1.16]. This so-called *packaging shift* can cause a significant temperature error, in the order of several tenths of a degree [2.28]. This shift can be corrected for by trimming after packaging, although the increased cost and thermal settling time associated with this are unattractive. Another option is to apply an anti-stress coating to the die, which also adds cost and may limit the operating temperature range.

The effects of mechanical stress on ETFs are unknown. Existing literature predicts that the thermal diffusivity of silicon is only a very weak function of mechanical stress [2.29], so D_{Si} is expected to remain unchanged. One way to measure the effect of stress is to package the same ETF (and its readout circuit) in both ceramic DIL (stress-free) and plastic TSSOP packages, the latter having the mechanical stress associated with plastic packaging. If warranted by the results of this experiment, a more quantitative measurement may be performed. Care has to be taken to limit the effect of self-heating, since differences in R_{th} will introduce self-heating errors. Chapter 5 will present some measurement results.

2.5.3 Resolution

The second aspect of ETF performance is their temperature-sensing resolution. As for most sensors, resolution is closely related to power consumption. This section will build and understanding of the sources of finite resolution in ETFs and their technology dependence. Expanding the modeling in this direction helps to design ETFs for optimal resolution, saving energy.

2.5.3.1 Establishing the resolution of silicon ETFs

The temperature-sensing resolution of an ETF is one of its most important parameters: because silicon is a good thermal conductor and since s may be large to achieve a target inaccuracy, a significant amount of heater power may have to be dissipated to generate a measurable output signal. In the earlier days of ETF-based temperature sensors, the required heater power was in the order of several tens of mW [2.22]. This compares unfavorably to state-of-the art BJT-based temperature sensors, which dissipate as little as 5µW and achieve 0.02ºC resolution at 5.3ms/conversion [2.24]. ETFs are fundamentally at a disadvantage since they have to generate heat to generate a temperature signal, so it is important to find out how this amount of heat can be minimized.

This section will derive the resolution of a point-heater, point-sensor ETF, then expand this to a basic thermopile with a hot and a cold junction and then further expand this to full ETFs.

As was shown in [2.23], the resolution of a simple ETF can be derived as follows. Consider a point heater, point sensor ETF driven by a square wave[2] having average power P_{heat}. Since the variable of interest is ϕ_{ETF}, the ETF is read out by a phase detector, implemented by a multiplier and a low-pass filter, together also known as a coherent detector. The multiplier is driven by a square wave having a first harmonic equal to $\cos(2\pi f_{drive}t + \phi_{demod})$. Fig. 2.32 shows the system:

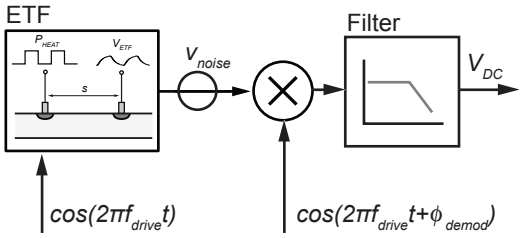


Fig. 2.32: Basic ETF interface for resolution analysis.

After coherent detection, the DC component of the filter output, V_{DC}, represents temperature information and is given by:

[2] This and further examples will use a square wave to drive the heater, since this is much easier to implement on chip as compared to e.g. generating a signal that generates sine-wave heat (P_{heat} is proportional to V_{heat}^2).

$$V_{DC}(s, f_{drive}, T) = \frac{8}{\pi^2} S_{tc} \frac{P_{heat}}{2\pi k_{Si} s} \exp(-s\sqrt{\frac{\pi f_{drive}}{D_{Si}}}) \cos(s\sqrt{\frac{\pi f_{drive}}{D_{Si}}} + \phi_{demod}) \tag{2.21}$$

As before, both the amplitude and the phase shift of this signal will be temperature-dependent.
The sensitivity of V_{DC} to temperature is now given by its derivative to T, which can be calculated by replacing k_{Si} and D_{Si} with their temperature dependencies. A further simplification lies in assuming that ϕ_{demod} will be controlled in such a way so as to drive V_{DC} to zero, i.e. the cosine term in Eq. (2.21) becomes $\cos(\pi/2)$. In this case, ϕ_{demod} will change over T and thus contain temperature information. This approximately maximizes $\partial V_{DC}/\partial T$, which is now given by [2.23]:

$$\frac{\partial V_{DC}(s, f_{drive}, T)}{\partial T} \approx 6.9 * 10^{-13} f_{drive}^{0.5} T^{1.2} \exp(-1.15 s f_{drive}^{0.5} T^{0.9}) \tag{2.22}$$

Fig. 2.32 also shows a noise source, v_{noise}. For now, assume that the input-referred noise of the phase detector is dominated by Johnson noise (=white noise) with a PSD of 4kTR (equivalent to about 4nV/√Hz for a 1kΩ resistor) and that the low-pass filter at the output limits the bandwidth to B. In this case, the temperature-sensing resolution is given by:

$$res = \frac{\sqrt{4kTRB}}{6.9 * 10^{-13} f_{drive}^{0.5} T^{1.2} \exp(-1.15 s f_{drive}^{0.5} T^{0.9})} \tag{2.23}$$

This equation was numerically evaluated for an ETF with an s of 24µm, driven at P_{heat}=2.5mW. A resistance of 1kΩ was used to determine the noise floor, although this value is somewhat arbitrary since the intrisinic noise of a point-heater, point-sensor ETF is poorly defined. However, it does define a reference level to study the f_{drive}- and temperature dependence of an ETF's resolution. The results for a 1Hz readout bandwidth are shown below:

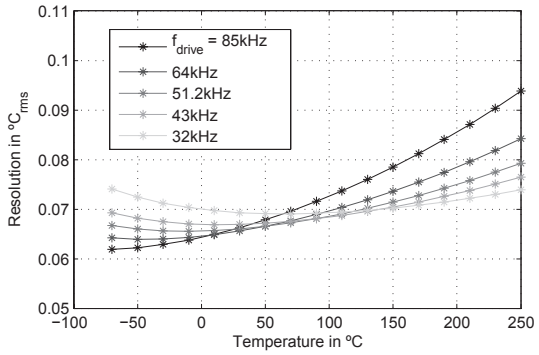


Fig. 2.33: Simulated resolution of an s=24µm point-heater point sensor ETF as a function of temperature and f_{drive}.

First, this model shows that large ETFs are indeed quite power-hungry: at 2.5mW, the moderate resolution of $0.06°C_{rms}$ is only achieved when the readout bandwidth is reduced to 1Hz, corresponding to a conversion rate of 2S/s. This underlines the importance of improving the resolution of ETFs.

Secondly, this model shows that resolution is a moderate function of f_{drive} and indicates that lower f_{drive} leads to lower noise, especially over temperature. However, the model above is still too basic to draw any hard conclusions.

In practice, the temperature sensor is implemented using thermocouples, and as such has both a hot and a cold junction, spaced at a distance L_{tp}, and the effective V_{ETF} is their difference (see Eq. (2.13)). Fig. 2.34 shows the system:

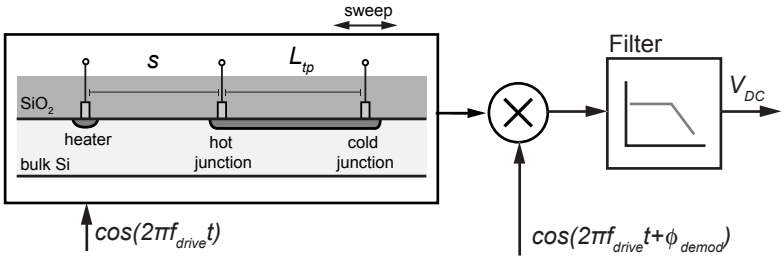


Fig. 2.34: ETF with single-arm thermopile sensor and phase detector.

L_{tp} represents a new degree of freedom. Intuitively, the cold junction should be far from the heater so that its signal is minimal. However, practical implementations of thermocouples in an IC substrate are made using diffusion resistors. Assuming a constant width, the thermal noise power of these resistors is a linear function of their length: increasing L_{tp} beyond some point will add more noise than remove signal, thus degrading resolution. This poses an interesting tradeoff that can be analyzed using the same thermal impedance model described in section 2.3.3. Additionally, there is a length L_{tp} for which the heat signal at the cold junction is exactly in anti-phase with that at the hot junction, increasing the effective V_{ETF}.

Fig. 2.35 shows the simulated ϕ_{ETF}, V_{ETF}, thermopile noise and resolution as a function of L_{tp}, for various f_{drive}. This is for an ETF with only one thermopile arm, having a width of 1µm and a sheet resistance of 96Ω/square (based on data from a 0.7µm CMOS process). As before, P_{heat}=2.5mW, s=24µm and B=1Hz, while L_{tp} was swept from 2 to 60µm. Temperature was kept constant at 27ºC.

Starting at the top left in a clockwise direction, Fig. 2.35 shows ϕ_{ETF}, V_{ETF} (peak-to-peak), resolution and thermopile thermal noise. ϕ_{ETF} clearly has a $\sqrt{f_{drive}}$ dependence, but also shows sensitivity to L_{tp}. As predicted, for very small L_{tp}, ϕ_{ETF} and V_{ETF} are reduced, thus reducing resolution. There is a weak optimum for L_{tp} that is dependent on L_{tp}, the peak of which is approximately at:

$$L_{tp,opt} \propto s / f_{drive}^{\,2} \tag{2.24}$$

This fits the intuitive expectation that lower frequency heat propagates further, and thus has a larger effect at the cold junction.

At and beyond $L_{tp,otp}$, the effect of L_{tp} on ϕ_{ETF} is relatively small. This corresponds to the results from section 2.5.2.2, where spread on the location of the cold junction was found to have a minimal effect.

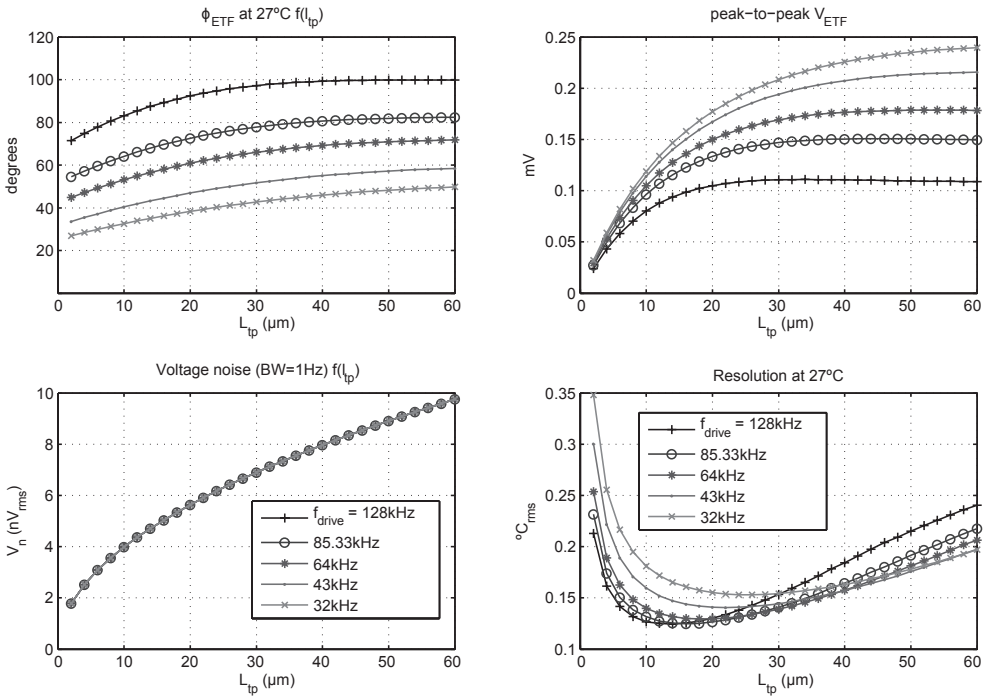


Fig. 2.35: Simulated ϕ_{ETF} , V_{ETF}, V_{noise} and resolution as a function of L_{tp}; s = 24µm, T=27°C, P_{heat} = 2.5mW; The ETF has a single thermopile arm.

The resolution of a complete ETF can be found by superposition of thermopile elements. When all thermopile arms have equal length and have their hot junction located on a phase contour (see section 2.3.3 on ETF modeling), their series-connected output has a signal-to-noise ratio given by:

$$SNR \propto \frac{n_{tc} S_{tc} (T_{hot} - T_{cold})}{\sqrt{4kTn_{tc}RB}} \quad , \tag{2.25}$$

in which n_{tc} is the total number of thermocouples and S_{tc} is their sensitivity. The equation shows that increasing n_{tc} leads to a square-root improvement in SNR. For best resolution, n_{tc} should always be maximized as much as area permits.

When s is given by accuracy requirements and L_{tp} is found from the simulation discussed above, the resolution of a complete ETF can now be simulated. The figure below shows resolution over T and f_{drive} for the ETF shown in Fig. 2.13, with s=24µm and an average L_{tp} of 22µm (phase-contour layout of the cold-junctions changes L_{tp} from arm to arm). There are 24 thermopile arms, each having a width of 2µm. R_{tp} is again based on data from a 0.7µm CMOS process, and the readout bandwidth is set to 0.33Hz. The temperature dependence of S_{tc} was not modeled. It can be seen that for the full ETF, resolution is weakly dependent on f_{drive} and T.

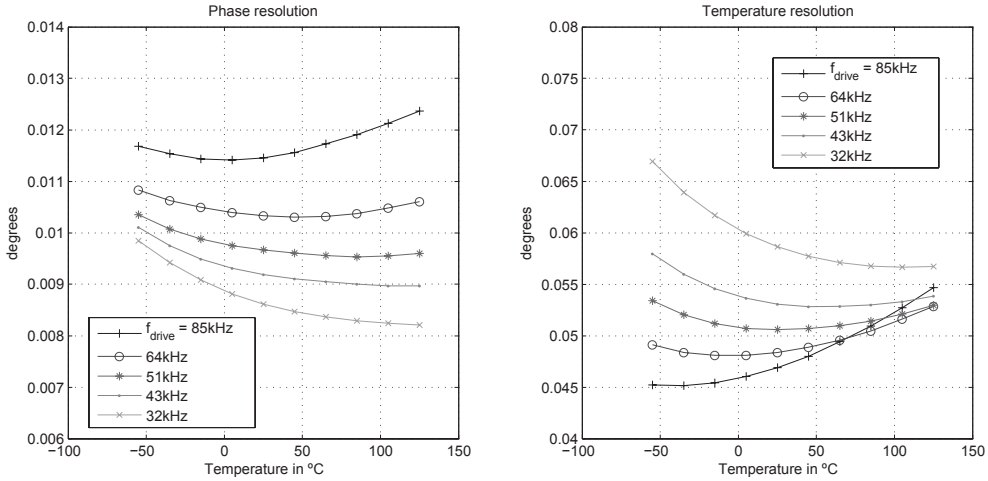


Fig. 2.36: Phase and Temperature resolution for an *s*=24µm ETF in a 0.7µm CMOS process.

The modeling assumes that P_{heat} is constant over temperature. In practice, this may be difficult to achieve since R_{heater} will be temperature-dependent. A P_{heat} control circuit could compensate for R_{heater} and possibly adjust P_{heat} for constant resolution over T.

2.5.3.2 Technology dependence and scaling

The main promise of ETF-based temperature sensors is their scaling behavior. While constant-size ETFs should have decreasing inaccuracy (see section 2.5.2.3), constant-error ETFs will have improved resolution in more advanced CMOS technology.

Intuitively, since the amplitude of the thermal signal increases sharply closer to the heater, resolution should improve by reducing *s*, i.e. scaling the ETF. However, scaling also linearly reduces $\phi_{ETF}(T)$, the signal that contains temperature information. The model discussed in the previous section can again be used to analyze the resolution of scaled ETFs.

Fig. 2.37 shows the resolution for a constant-error ETF with *s*=24µm in 0.7µm CMOS, which thus has *s*=12µm in 0.35µm technology and proportionally smaller *s* for further scaling. The resolution is calculated in a 1Hz bandwidth. The slope of the logarithmic plot is about 1.5, indicating that resolution improves with $s^{-1.5}$ (as predicted in [2.23]).

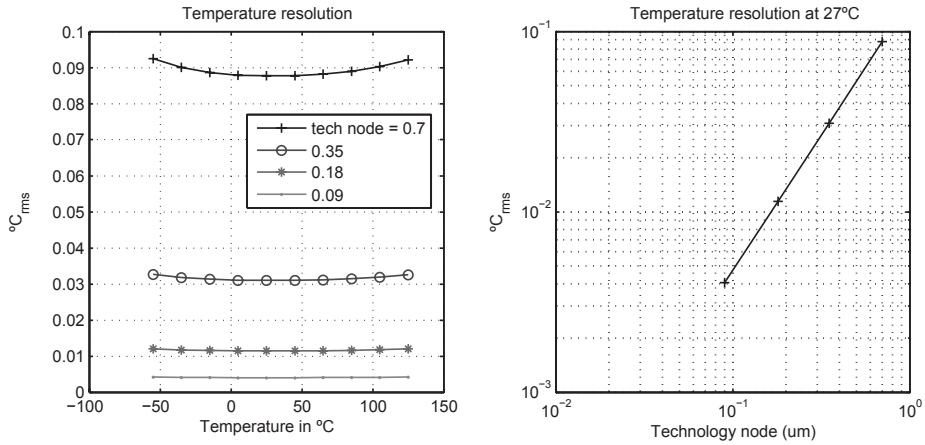


Fig. 2.37: Constant-error ETF resolution as a function of technology node.

It is also of interest to observe how the resolution optimum for different values of s depends on f_{drive}. From Equation (2.24), the optimum f_{drive} will increase strongly for decreasing s. This is illustrated by Fig. 2.38:

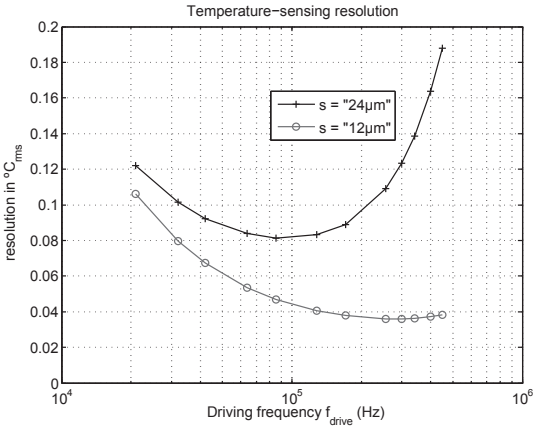


Fig. 2.38: Resolution versus f_{drive} for an s=12 and an s=24µm ETF.

While the above results show that resolution improves strongly with scaling, it is of interest to see what amount of improvement is required for ETFs to become competitive to BJTs for microprocessor thermal management, which is, as was argued in Chapter 1, the field in which scaled ETFs are expected to be most effective.

Typical requirements for thermal management are P_{heat} = 1mW, a resolution of 0.1°C$_{rms}$, an absolute accuracy of ±1°C (3σ) and a conversion rate of about 1kS/s. This corresponds to a bandwidth of 500Hz, which is much higher than has been used so far: strong scaling improvements are required to achieve such resolution.

Fig. 2.39 shows conversion rate as a function of technology node for three different ETFs. The lower curve corresponds to the ETF that has been used to far, with a lithography-defined spread of ±0.5°C (3σ). The second curve corresponds to an ETF for which this budget has been increased to ±1.0°C, thus halving s and improving resolution. The top curve corresponds to a ±0.5°C ETF in SOI technology, which, as will be shown shortly, offers better resolution yet.

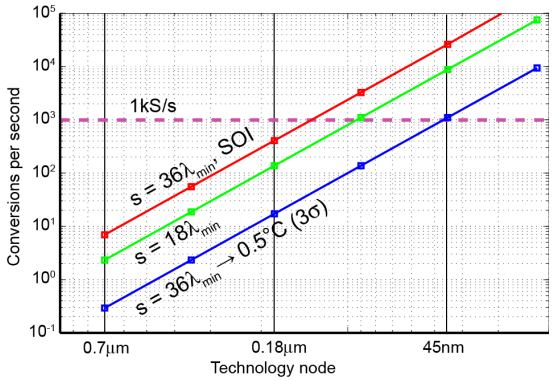


Fig. 2.39: Predicted SNR for ETFs in scaled CMOS technology.

Based on these modeling results, ETFs in deep submicron CMOS should be able to easily achieve the high conversion rates required for thermal management.

While most resolution improvement can be expected from scaling, the intrinsic resolution of thermopiles also improves in more advanced CMOS technology. As source and drain diffusions become more heavily doped, their sheet resistance goes down, improving noise. The Seebeck coefficient (S_{tc}) is also process-dependent. Typically reported to be about 0.5mV/K in older technologies [2.13], this can be significantly higher in modern processes, because S_{tc} for an aluminum-diffusion resistor is proportional to ln ρ/ ρ. In modern processes with lower ρ, this will increase S_{tc}. Lastly, the decreasing minimum spacing between thermopile arms allows more area around the ETF to be used as thermopile, thus reducing R_{tp} further. These three mechanisms all further improve resolution in scaled CMOS technology.

2.5.3.3 Moving to SOI technology

Silicon on insulator (SOI) technology is used for high voltage, high precision circuits, but also increasingly in advanced deep submicron technologies. In SOI technology, a relatively thick SiO_2 layer isolates the substrate from the epi-layer, which can be patterned by isolation trenches to electrically isolate circuits from each other. SiO_2 isolates electrically, but also thermally, and so it is of interest to compare ETFs in SOI to their bulk CMOS counterparts.

Intuitively, it is to be expected that the heater of a silicon ETF in SOI will see a greater thermal impedance, because the buried oxide (BOX) layer will prevent heat from diffusing into the substrate. FEM simulations predict [2.17] that the associated increase in amplitude at the thermopile is about 5x, which would directly translate into improved SNR or P_{heat} reduction.

In order to extend the point-sensor, point-heater analytical ETF model to SOI, the effective thermal impedance has been analyzed by using the mirror image technique [2.17], also used in analyzing BJT self-heating in SOI [2.18]. To model the effect of a thermally insulating layer, heater mirror images are added to ensure that the effective flow through the BOX layer is zero. The silicon surface is also considered a thermal insulator, requiring extra mirror heaters. The signals of these mirror images add up and increase dT at the sensor, although these mirror images are at progressively longer effective distances from s, so that their magnitude and phase characteristics are different. In practice, since the magnitude of Z_{th} rapidly falls off for increasing r, only mirror image heaters with $\sim r_{,eff} < 2r_1$ contribute significantly (r_1 is the radius of the initial (i.e non-mirrored) heater-sensor structure).

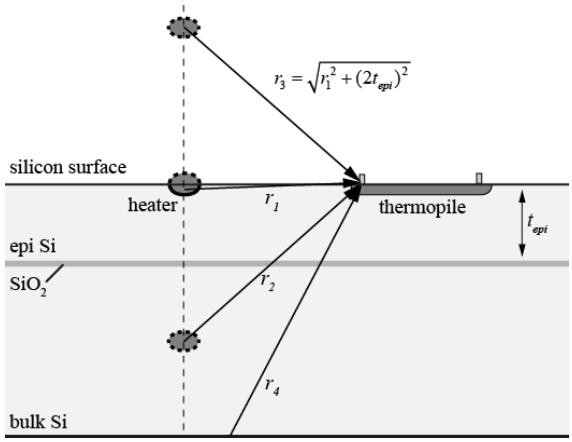


Fig. 2.40: Modeling technique to include the effect of a buried oxide layer on ϕ_{ETF}.

Using this analytical model, changes in ϕ_{ETF} as a function of the depth of the buried oxide layer can be calculated.

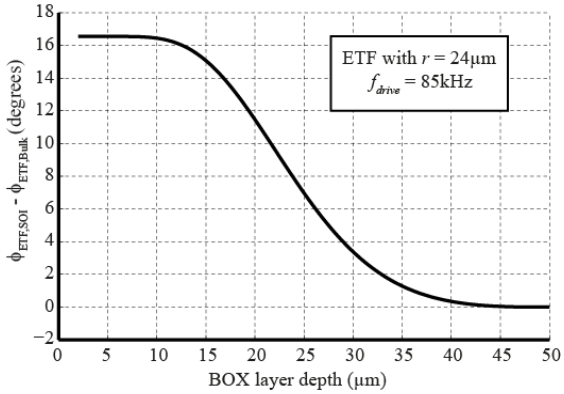


Fig. 2.41: Additional phase shift as a function of BOX layer depth.

Fig. 2.41 shows the simulated offset as a function of the BOX layer depth; $\phi_{ETF,SOI}$ clearly has a positive offset to $\phi_{ETF,BULK}$. Although this is in line with measurements (see chapter 5), the analytical model overestimates the signal amplitude and does not include the effects of spread in BOX layer thickness. In practice, some heat also flows through the SiO_2 layers. However, taking this into account makes analytical modeling prohibitively complex. Moreover, the underlying assumption that heat diffuses through a semi-infinite medium no longer holds. Finite element simulations are more accurate, but require much more computational effort. Further work is required to develop accurate analytical models for ETFs in SOI.

The measurements section will discuss SNR and accuracy measurements on ETFs in Bulk CMOS and SOI CMOS technology.

2.5.3.4 Energy per conversion

To further understand ETF performance and enable better comparison to other temperature sensors, it is of interest to calculate the energy per conversion for an ETF-based temperature sensor.

Assuming that SNR is limited by thermal noise in the ETF, and that the power consumed in the readout circuit ($P_{circuit}$) is significantly smaller that P_{heat}, the energy per conversion can be approximated by:

$$E_{conv} \approx P_{heat} * t_{conv} \tag{2.26}$$

For the large ETFs discussed above, E_{conv} is on the order of several mJ, which is several orders of magnitude larger than BJT sensors.

ETF-based sensors are different from most sensors in the sense that a linear increase in their power consumption leads to a quadratic improvement signal improvement, since V_{ETF} is linearly proportional to P_{heat}:

$$SNR \propto \frac{V_{ETF}}{\sqrt{4kTRB}} \propto P_{heat} * \sqrt{t_{conv}} \tag{2.27}$$

At constant SNR, lowest energy per conversion is thus achieved by maximizing P_{heat} (thereby reducing t_{conv}). Consequently, it is most energy-efficient to operate the ETF for the shortest possible time required to achieve a target resolution, extending to duty-cycle operation of ETFs.

In practice, there are several limits to P_{heat}:

- ETFs will locally self-heat, adding a temperature offset to T_{die}. While this could be calibrated out, the spread in this self-heating will introduce device-to-device spread; this is a linear function of P_{heat}.
- At and beyond a certain P_{heat}, other noise sources such as quantization noise may limit the resolution, at which point Eq. (2.27) no longer holds.
- Especially at high temperatures and/or low supply voltages, electromigration will limit the maximum P_{heat} that can be generated.

In the measurement section, SNR as a function of f_{drive}, s and process technology will be discussed.

2.5.3.5 Alternative solutions

A more exotic way of improving an ETF's SNR is that of etching away silicon to create a very small thermal mass in the area of the ETF, which will now be located in a thin silicon bridge (Fig. 2.42). While this potentially offers order-of-magnitude improvement, it is prohibitively expensive for most applications.

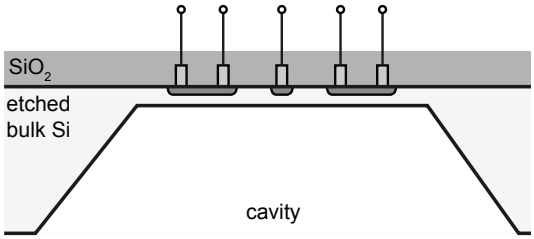


Fig. 2.42: ETF implemented on a thermally isolated silicon bridge.

Interestingly though, the bridge ETF will also be sensitive to the thermal diffusivity of the material in the cavity, which is potentially useful for e.g. gas sensors [2.20]. However, in the context of this thesis, this direction is not explored further: in thermal management applications, etching away silicon will impede heat flow and thus increase self-heating, which is undesirable.

2.5.4 Temperature Range

As Fig. 2.5 shows, both D_{Si} and D_{SiO2} are well-defined over an extremely wide temperature range. ETF-based temperature sensors thus offer the advantage of being able to work at extreme temperatures, for example in space or industrial applications (such as down-hole sensing in oil drilling). It is of interest to study the widest temperature range over which ETF-based sensors can be made to operate.

At very high temperatures, the SNR of silicon ETFs will deteriorate both by increasing thermal noise and reducing sensitivity (D_{Si} flattens out at very high T), but both can be alleviated by increasing P_{heat}. More fundamental limitations are likely to be associated with reliability issues such as electromigration, although the operational lifetime in e.g. downhole drilling applications is typically limited to several hours only.

At very low temperatures, λ_{ph} increases as phonon energy decreases and phonon-phonon scattering becomes less dominant. Therefore, k_{Si} and D_{Si} are more strongly affected by phonon-boundary scattering and phonon-impurity scattering. Lattice defects or dopant atoms will reduce D_{Si} from its ideal value, and phonon-boundary scattering on the edges of the die can also become problematic [2.6]. Fortunately, due to increasing dD_{Si}/dT, the corresponding temperature error is limited.

Wide-temperature range operation of ETFs can only be shown if their readout circuitry is also functional and accurate over temperature. Chapter 4 will discuss readout circuitry that is robust to e.g. high temperature leakage currents and varying threshold voltages and sheet resistances.

Lastly, the test setup and reference temperature sensor both should not limit the range. Chapter 5 will present wide-range measurements that provide a quantified answer to the above concerns.

2.5.5 Size

The last performance aspect to be discussed is sensor size, or more concretely, the amount of silicon die area needed to co-integrate the temperature sensor with other circuitry on the chip. This is essential for thermal management applications, since many temperature sensors may be spread out across the chip, while their total area should ideally still be a small fraction of the die size.

It is clear that the area of an ETF is strongly determined by the heater-sensor spacing, s. The area of an ETF can be estimated by:

$$area = 12s^2 \tag{2.28}$$

This equation approximates the area of a phase-contour ETF with $L_{tp} \approx 2s$, which is a typical value from a resolution standpoint.

ETFs with large s (e.g. 24µm) in older 0.7µm technologies will be quite large, at ≈7000µm^2. A typical bipolar core, using e.g. an 8:1 configuration of 5x5µm PNP devices, is about 30x smaller. However, nanometer CMOS technology allows for much smaller ETFs, and an s = 1µm ETF will be of similar area to a BJT.

2.6 Concluding remarks

This chapter has provided some theoretical background, modeling and design strategies for ETFs. It has also provided an overview of some of the important aspects of ETF performance that should be studied to evaluate whether ETFs can be competitive integrated temperature sensors. Chapter 5 will present measurement results that relate to each of these aspects, in an attempt to validate the predictions given above. Before moving on to measurement results though, Chapters 3 and 4 will first discuss ETF interfacing architectures and readout circuits.

2.7 References

[2.1] R. Berman, “Thermal Conduction in Solids”, Oxford University Press, 1976.

[2.2] M. Kaviany, "Essentials of Heat Transfer", Cambridge University Press, 2011.

[2.3] H.S. Carslaw and J.C. Jaeger, “Conduction of Heat in Solids”, Oxford University Press, 1959.

[2.4] L. Lou, “Introduction to Phonons and Electrons”, World Scientific Publishing, 2003.

[2.5] D.V. Schroeder, “Thermal Physics”, Addison Wesley Longman, 2000.

[2.6] M. G. Holland, “Analysis of Lattice Thermal Conductivity”, Physical Review, vol. 132, no. 6, pp. 2461-2471, Dec. 1963

[2.7] M. Asheghi et al., “Thermal conduction in doped single-crystal silicon films”, Journal of Applied Physics, Vol. 91, pp. 5079-5088, Apr. 2002

[2.8] P. Turkes, “An ion-implanted resistor as thermal transient sensor for the determination of the thermal diffusivity in silicon,” Phys. Status Solidi A, vol. 75, no. 2, pp. 519-523, 1983.

[2.9] H.R. Shanks et al., “Thermal conductivity of silicon from 300 to 1400K”, Phys. Rev. (USA), Vol. 130, pp. 1743 – 1748, 1963.

[2.10] M. Asheghi et al., “Temperature-Dependent Thermal Conductivity of Single-Crystal Silicon Layers in SOI Substrates”, Trans. ASME, vol. 120, pp. 30 – 36, Feb. 1998.

[2.11] M. B. Kleiner, S.A. Kühn and W. Weber, “Thermal Conductivity of Thin Silicon Dioxide Films in Integrated Circuits”, Proc. ESSDERC 1995, pp. 473-476, Sept. 1995.

[2.12] Y.S. Touloukian et al., Thermophysical Properties of Matter: Vol. 10. New York: Plenum, 1998.

[2.13] S. van Herwaarden, “Thermal Vacuum Sensors based on Integrated Silicon Thermopiles”, Ph.D. Thesis, Delft University of Technology, 1987.

[2.14] T. Veijola, “Simple model for thermal spreading impedance”, Proc. BEC’96, pp. 73-76, Oct. 2005

[2.15] S. Xia, K.A.A. Makinwa, “Design of an optimized electrothermal filter for a temperature-to-frequency converter”, Proc. IEEE Sensors 2007, pp. 1255 – 1258, Oct. 2007.

[2.16] S.M. Kashimiri and K.A.A. Makinwa, “Measuring the Thermal Diffusivity of CMOS chips”, proc. IEEE Sensors, pp. 45-58, November 2009.

[2.17] S. Xia, “Modeling and Optimization of an Electrothermal Filter”, M.Sc. Thesis, Delft University of Technology, Delft, The Netherlands, July 2007.

[2.18] A. Pacelli, P. Palestri and M. Mastrapasqua, “Compact Modeling of Thermal Resistance in Bipolar Transistors on Bulk and SOI substrates”, IEEE Trans. Electron Devices, vol. 49, no.6, pp. 1027-1033, June 2002.

[2.19] B. Vermeersch, “Thermal AC Modelling, Simulation and Experimental Analysis of Microelectronic Structures Including Nanoscale and High-speed Effects”, Ph.D. Thesis, Ghent University, Ghent, Belgium, 2009.

[2.20] R.P. Manginell, J. H. Smith, A. J. Ricco, “An overview of micromachined platforms for thermal sensing and gas detection”, Proceedings of SPIE on International Society of Optical Engineering, 1997, pp. 273-284.

[2.21] V. Szekely, “Thermal monitoring of microelectronic structures”, Microelectronics J., vol. 25, no. 3, pp. 157–170, 1994.

[2.22] V. Szekely and M. Rencz, “A new monolithic temperature sensor: the thermal feedback oscillator,” Digest Transducers, Jun. 1995, vol. 15, pp. 849–852.

[2.23] S.H. Shalmany, “Thermal Diffusivity-Based Temperature Sensors for Microprocessors”, M.Sc. Thesis, Delft University of Technology, Delft, The Netherlands, Oct. 2010.

[2.24] K. Souri, Y. Chae and K.A.A. Makinwa, “A CMOS Temperature Sensor With a Voltage-Calibrated Inaccuracy of ±0.15°C (3σ) From -55 to 125°C,” Digest ISSCC, pp. 208 – 210, Feb. 2012.

[2.25] K.A.A. Makinwa and J.F. Witte, “A Temperature Sensor based on a Thermal Oscillator,” Proc. IEEE Sensors, pp. 1149 – 1152, Oct. 2005

[2.26] K.A.A. Makinwa and M.F. Snoeij, "A CMOS temperature-to-frequency converter with an inaccuracy of ±0.5°C (3s) from –40 to 105°C," J. Solid-State Circuits, vol. 41, is. 12, pp. 2992 – 2997, Dec. 2006

[2.27] J. Creemer, F. Fruett, G. Meijer, and P. French, "The piezojunction effect in silicon sensors and circuits and its relation to piezoresistance," IEEE Sensors J., vol. 1, no. 2, pp. 98 –108, Aug. 2001.

[2.28] 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.

[2.29] K. Alfatooni and A. Nathan, "Heat transport properties of semiconductors under nonuniform stress", Appl. Phys. Lett., Vol. 66, No.9 , Feb. 1995.

[2.30] Q. Qian et al., "The optimization of deep trench isolation structure for high voltage devices on SOI substrate", J. Solid-State Electronics, Vol. 63, No. 1, pp. 154-157, Sept. 2011.

3 ETF Systems

3.1 Introduction

Having presented the theory and operating principles of ETFs in Chapter 2, this chapter will discuss how ETFs can be read out, and the various systems that can be built around them. Based on the requirements of typical applications, the architecture of such systems and the requirements on their various building blocks will be discussed.

As shown in Chapter 2, the relationship between the driving frequency f_{drive}, temperature T and resulting phase shift ϕ_{ETF} of both silicon and oxide ETFs can be expressed as:

$$\phi_{ETF} \propto \sqrt{f_{drive} T^C} \tag{3.1}$$

In which C is a material-dependent constant ($C \approx 1.8$ for Si, $C \approx 0$ for SiO_2). In the context of on-chip components, this relationship provides a new link between these three variables, which can be exploited to realize various ETF-based systems, several of which are identified in Table 3.1. Each system is a combination of several building blocks, each performing a fundamental operation such as *measuring*, *adjusting*, and/or maintaining *constant* one of the three variables: phase, frequency or temperature. Assuming, for now, that each of these building blocks is feasible and implementable, this chapter will discuss four classes of ETF systems:

1. **ETF-based temperature-to-frequency converters** (section 3.2). These regulate the ϕ_{ETF} of a silicon ETF to a *constant* value by *adjusting* f_{drive}. Their output frequency is a function of absolute temperature, which can be digitized by a frequency-to-digital converter.
2. **ETF-based temperature-to-digital converters** (section 3.3), These drive a silicon ETF with a *constant* f_{drive} and *measure* ϕ_{ETF}. As a result, ϕ_{ETF} will be a function of absolute temperature, which can then be digitized by a phase-to-digital converter.
3. **ETF-based frequency references.** These *measure* temperature (with another temperature sensor) and then *adjust* ϕ_{ETF} accordingly so as to regulate f_{drive} to a target *constant* value. Although not the main focus of this work, their basic operation will be briefly discussed in section 3.4.
4. **ETF-based self-referenced temperature sensors.** These use two ETFs driven at a *constant* frequency; they *measure* the ratio of the phase shifts of a silicon and an oxide ETF. Such sensors do not require an external time reference (section 3.5).

Table 3.1: Overview of several ETF-based systems

ETFs	Constant	Adjust	Measure	Output
D_{Si}	Phase (e.g. PLL)	Phase (controlled phase shifter)	Phase (phase detector)	T sensor (section 3.3)
D_{SiO2}	Frequency (e.g. xtal)	Frequency (VCO, DCO)	Frequency (e.g. counter)	T to freq. converter (section 3.2)
D_{Si} & D_{SiO2}	Temperature (oven)	Temperature (Oven)	Temperature (temp. sensor)	frequency reference (section 3.4)
			Phase Ratio (phase detector)	ratiometric temp. sensor (section 3.5)

The creative mind will recognize that the above list is not exhaustive; other ETF-based systems can be constructed and other functions defined. Some alternatives, such as those requiring active temperature control, are probably impractical and/or outside the scope of this work. Some of the more promising alternatives will be briefly discussed in the Future Work section (Ch. 6).

This chapter will continue by presenting each of the four systems outlined above, discussing the functionality and performance requirements for each of these blocks in more detail. Although silicon ETFs can be expected to have low untrimmed inaccuracy, limited doping- and stress sensitivity and a wide operating range, they typically output small, noisy AC signals. To process such signals accurately, ETF-based systems will typically employ coherent detection and narrowband filtering, which mitigate the effects of 1/*f* noise and thermal noise, respectively.

Finally, since integrated temperature sensors typically have a digital output, methods of accurately *digitizing* a thermal delay are discussed. Since the variable of interest is in the time domain, this will require frequency-, phase-, or delay-domain ADCs. It will be shown that phase-domain sigma-delta modulators (PDΣΔMs) are highly suited to the read-out of ETFs with high accuracy and low noise. Since these converters operate in the time domain, their precision can be expected to improve significantly in faster, more advanced CMOS technology.

3.2 Temperature-to-Frequency converters

3.2.1 General concept

The first implementation of a temperature sensor based on sensing the temperature dependence of D_{Si} is most likely the work of Szekely et al. [3.1][3.2]. This work describes a "Thermal Feedback Oscillator" (TFO), i.e. a temperature-to-frequency converter of which the output frequency is a function of D_{Si}, and thus of absolute temperature.

The TFO's operation is based on an ETF that is more complex than those described in Chapter 2. Rather than implementing a low-pass filter, a bandpass-filter was implemented by combining the outputs of several temperature sensors along the heat diffusion path so as to create a step response whose zero crossing is defined by geometry and D_{Si}. This is illustrated in Fig. 3.1:

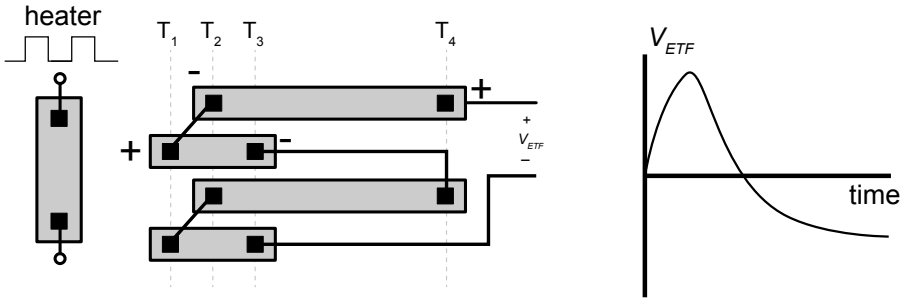


Fig. 3.1: Simplified drawing of the ETF used in the Thermal Feedback Oscillator (TFO).

This ETF combines the signal at four temperatures, $T_1 - T_4$, in the polarity shown above to obtain the step response plotted on the right side of Fig. 3.1. As a result, a simple relaxation oscillator can be built in which a comparator (positively) feeds back an amplified version of the ETF's output to its heater [3.2]. Such a system will oscillate at the frequency where the thermal delay corresponds to a phase shift of 180 degrees, while driving the average ETF output voltage to zero. While it was shown that the output frequency of such an oscillator is indeed temperature-dependent, its accuracy and sensitivity to process spread were not investigated. Moreover, the oscillator suffered from poor jitter, due to the ETF's sub-millivolt output signal and the absence of any noise filtering before the comparator (see e.g. [3.3]).

An improved method of locking an oscillator to the thermal delay of a silicon ETF was pioneered by Witte and Makinwa [3.4]. Realizing that the ETF's small output signal can only be read out with high resolution in a narrow bandwidth, they embedded a low-pass ETF in a frequency-locked loop (FLL), such as the one shown in Fig. 3.2:

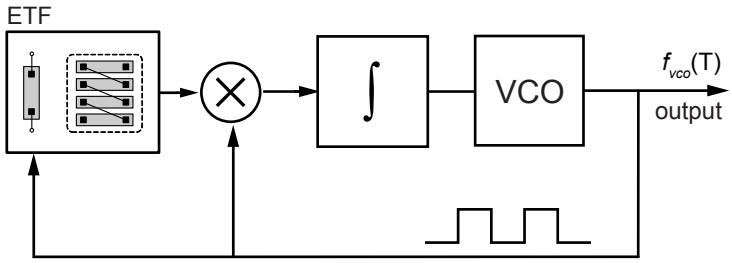


Fig. 3.2: Analog Frequency-locked loop (FLL)

In this loop, an integrator drives a VCO such that it outputs a frequency (f_{vco}) for which the average integrator input equals zero. A multiplier performs phase detection by multiplying f_{vco} with the ETF output signal, which is a low-pass filtered version of f_{vco}. When only the first harmonic of the f_{vco} square wave is considered, the DC component of the multiplier output is given by:

$$V_{out,DC} = \int_0^{1/f_{vco}} \cos(2\pi f_{vco} t)\cos(2\pi f_{vco} t + \phi_{ETF})dt = \frac{1}{2}\cos(\phi_{ETF}) \tag{3.1}$$

An ideal integrator will force $V_{out,DC}$ to zero, thus forcing ϕ_{ETF} to 90º. This performs one of the key functions shown in Table 3.1, i.e. that of maintaining a constant phase shift. As a result, f_{vco} should be proportional to $1/T^{1.8}$, as was confirmed by measurements [3.5].

The combination of a phase detector and an integrator acts as a narrowband filter. This reduces the FLL's noise bandwidth, thus suppressing the wideband thermal noise from the ETF's thermopile and increasing the system's resolution at the cost of bandwidth / response time. Using a linear model for both ETF and VCO gains and an integrator having a unity-gain bandwidth f_{ugbw}, the loop bandwidth is given by:

$$BW = K_{ETF} f_{ugbw} K_{VCO} \tag{3.2}$$

An analog FLL with an s=20µm D_{Si} ETF driven at P_{heat} =2.5mW has a temperature-sensing resolution of about 0.04ºC$_{rms}$ for a loop bandwidth of 0.5Hz [3.5]. For classical RC or g_m-C integrator architectures, such low bandwidths will require large capacitors (e.g. 1µF in [3.5]) that cannot be integrated on-chip. While the relatively poor SNR is intrinsic to ETFs with large s, the need for external passives is a particular disadvantage of the analog FLL.

3.2.2 Design considerations

The temperature-sensing inaccuracy of an FLL is determined by errors in the ETF (see Ch.2) and by errors in the readout circuit. When the ETF is designed to have a ±0.5ºC (3σ) inaccuracy so as to meet the requirements of typical commercial and industrial temperature sensing applications, circuit errors should ideally be at the ±0.05ºC level. The design of such precision circuits will be discussed in more detail in Chapter 4, but for completeness, several error sources are briefly discussed below:

- Timing errors in the demodulator will introduce a phase error (ϕ_e) in ϕ_{ETF}. These will be small enough if the demodulating circuitry and its associated logic are fast enough. The phase error corresponding to 0.05ºC error is given by $\approx 2s\sqrt{f_{drive}}$, or down to about 10 mº phase for the s = 20µm example.
- The offset and drift of the integrator also cause steady-state errors. These can be reduced through the use of dynamic offset-cancellation techniques such as chopping and auto-zeroing [3.12].
- The DC gain of the loop needs to be high enough to make the static gain error sufficiently small. More quantitatively, < 0.01ºC error requires a DC gain of at least 85dB, which typically requires the use of multiple gain stages. The reader is referred to [3.6] and [3.7] for examples.
- Errors in the VCO; these are suppressed by the DC gain of the loop.

Later implementations of the FLL include a digital filter in the loop. A block diagram of such a 'digitally assisted' FLL (DAFLL, from [3.8]) is shown below. The phase detector is implemented by a phase-to-digital converter, followed by a phase subtraction operation in the digital domain.

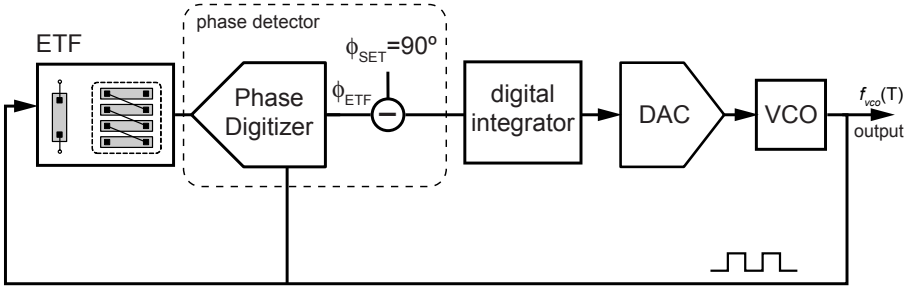


Fig. 3.3: Digitally assisted FLL (DAFLL).

By digitizing ϕ_{ETF} before phase detection and integration, the dominant time constant can be realized in the digital domain, thus no longer requiring large (off-chip) passives in the analog front-end while also implementing an infinite DC gain. A digitally-controlled oscillator (DCO) must now be used to generate f_{drive}, which can be implemented e.g. by a current DAC driving a VCO as in [3.8]. The DAFLL requires a high-resolution phase digitizer; this building block will be discussed shortly.

3.2.3 Typical performance

Some measurements results from FLLs based on an s=20µm silicon ETF in a 0.7µm CMOS process are presented below. As expected from theory, f_{vco} follows a $1/T^{1.8}$ temperature proportionality and the device-to-device spread in frequency is on the order of ±0.25% (3σ) [3.5][3.7]. Most of this spread can be attributed to lithographic spread in the ETF.

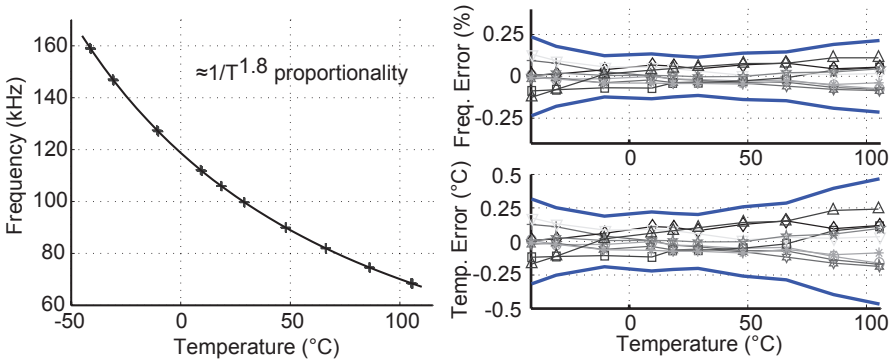


Fig. 3.4: Measured performance of D_{Si} ETF-based FLL.

Since f_{vco} is not a dimensionless parameter, measuring absolute temperature using an FLL involves comparing f_{vco} to a (temperature-independent) reference frequency, such as can be generated by a crystal oscillator. A microcontroller can then compute T_{abs} from a frequency *ratio*, and the resulting temperature sensing inaccuracy is also shown in Fig. 3.4 [3.5][3.7].

Although an untrimmed inaccuracy of ±0.5ºC (3σ) is comparable to the most accurate batch-calibrated BJT-based temperature sensors, f_{vco} is substantially nonlinear; its $1/T^{1.8}$ temperature dependence is a fundamental disadvantage of FLLs.

Frequency-output ETF systems are potentially useful for remote temperature sensing in noisy and harsh environments, since they use a single wire to output a large FM signal that is insensitive to noise and duty-cycle variations. However, a system that simultaneously reads out multiple ETFs in close proximity by using multiple FLLs may be prone to effects such as injection locking. In such cases (e.g. for µP thermal management), a more robust approach is to use a temperature-to-digital converter to directly digitize the thermal delay of an ETF. This also simplifies communication with a µP, since standard digital interfaces can be used.

3.3 Temperature-to-Digital converters based on a single D_{Si} ETF

3.3.1 General concept

Digitizing an ETF's thermal delay requires comparison to a reference delay. By integrating this function into the readout circuit, a simple and efficient temperature-to-digital converter can be made. As in section 3.2, a crystal oscillator can serve as the time reference.

The reference comparison can be done in two ways, i.e. either by comparing $\phi_{ETF}(T)$ to a reference phase ϕ_{REF}, or by comparing $f_{vco}(T)$ to a reference frequency, f_{REF}. The latter follows naturally from the FLL and will be discussed first.

Fig. 3.5 shows a system similar to the DAFLL, in which the combination of a DAC and VCO is replaced by a tunable frequency divider with ratio N, dividing down f_{ref}. In steady state, the digital integrator will output the digital code for which $f_{drive} = f_{ref}/N$ most closely corresponds to the frequency required to obtain a 90º phase shift between the ETF output and f_{drive} (and thus zero DC input to the digital integrator). The resulting digital output N is a function of T_{abs} that will again have a $1/T^{1.8}$ temperature dependence.

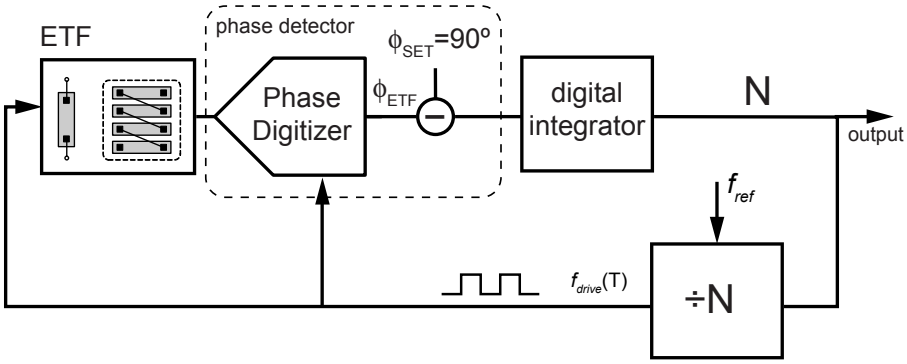


Fig. 3.5: Frequency-feedback thermal delay digitizer.

The precision and resolution of this system are defined by error sources similar to those in the FLL, as well as two new error sources associated with the time-to-digital conversion. Firstly, any error in f_{ref} (including e.g. its temperature dependence) will lead to an error in the digital output. Through Eq. (3.1), it can be shown that a temperature error of ±0.05ºC requires f_{ref} to have a maximum absolute error of ±225ppm (at worst case, T=125ºC). Secondly, achieving a 0.05ºC resolution over a -55 to 125ºC range requires at least a 12-bit frequency divider.

For a maximum f_{drive}(T) of 200kHz (at -55ºC, extrapolated from Fig. 3.4), a non-fractional frequency divider would require an f_{ref} of $f_{drive} * 2^{12} \approx$ 800MHz, with ±225ppm accuracy. Such high-speed circuitry would require careful design and significant power consumption. This is especially unattractive for arrays in which multiple ETFs are interfaced simultaneously (e.g. in thermal management applications).

3.3.2 Oversampled frequency readout

Fortunately, since temperature is typically a slow signal, oversampling and noise shaping techniques can be used to reduce the required f_{ref}. Because the above loop has integrating properties, it will still, over time, output a fractional N value even when the divider does not have high resolution (assuming a high-resolution phase digitizer). In the limit, such a system requires only a single-bit frequency DAC; the Fig. 3.5 loop can thus be converted to a first-order, single bit frequency-domain sigma delta modulator (FDΣΔM) [3.10][3.11].

The oversampled, single-bit output *bitstream* (bs) of a FDΣΔM can be low-pass filtered by a decimation filter to achieve the required resolution, effectively computing a fractional N value. The decimation filter digitally determines the output bandwidth and is outside the loop, so the FDΣΔM does not require a very low bandwidth loop filter; a simple analog integrator that integrates the comparator's quantization error is sufficient. The integrator typically requires capacitors in the order of several tens of pF, which can easily be integrated on-chip. Fig. 3.6 shows the resulting FDΣΔM and typical timing waveforms:

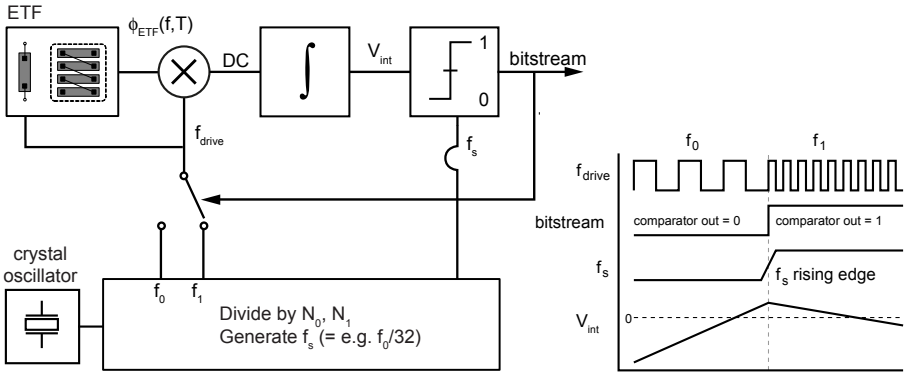


Fig. 3.6: First-order, frequency-domain sigma-delta modulator (FDΣΔM).

The output bitstream will represent a weighted average of two reference frequencies (f_0 and f_1) that straddle the f_{vco}(T) range. f_0 and f_1 can be straightforwardly generated by dividing down f_{ref}, whose value can now be much lower (i.e. as low as f_1). One constraint on the choice of frequencies is that the modulator's sampling period should be an integer number of f_0 and f_1 periods.

While the FDΣΔM provides an elegant way of digitizing ϕ_{ETF}, its output is still rather non-linear. The temperature dependence of the bitstream average will be similar to that of the FLL, i.e. $1/T^{1.8}$. Additionally, since the output of the phase detector is not exactly zero (only *on average*), a cosine nonlinearity is introduced in the FDΣΔM transfer function (discussed further in Chapter 4). Also, since the ETF's heater is switching between widely different f_{drive}'s while some of its thermal time constants are lower than f_{drive}, inter-symbol-interference (ISI) may occur.

3.3.3 Oversampled phase readout

3.3.3.1 Architecture development

A more linear way of converting a D_{Si} ETF's thermal delay into absolute temperature is to operate in the phase domain, noting that at constant f_{drive}, ϕ_{ETF} is proportional to $T^{0.9}$. A phase-to-digital converter having a digital output of the form ϕ_{ETF}/ϕ_{REF} is conceptually shown in Fig. 3.7. Since the ETF is driven by a constant-frequency bitstream-independent signal, its thermal-domain dynamics are periodic and so no thermal ISI can occur. This advantage and the greatly improved linearity make phase-domain ETF readout architectures preferable to their frequency-domain counterparts.

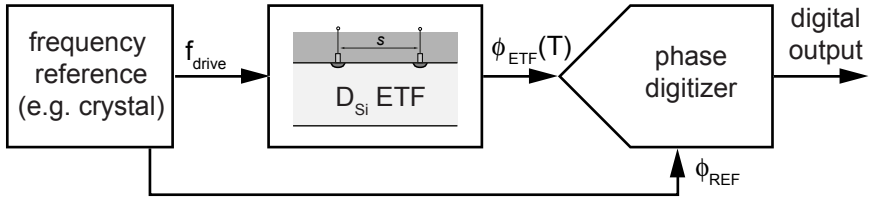


Fig. 3.7: ETF-based temperature sensor employing phase readout.

The Fig. 3.6 FDΣΔM can easily be transposed into a phase-domain sigma-delta modulator (PDΣΔM), as shown in Fig. 3.8. Instead of feeding back two frequencies, the bitstream now drives a phase DAC that selects one of two phase shifts, ϕ_0 and ϕ_1, that span ϕ_{ETF}(T). These phase shifts are generated e.g. by using a digital counter to synchronously delay f_{drive} by periods of f_{ref}.

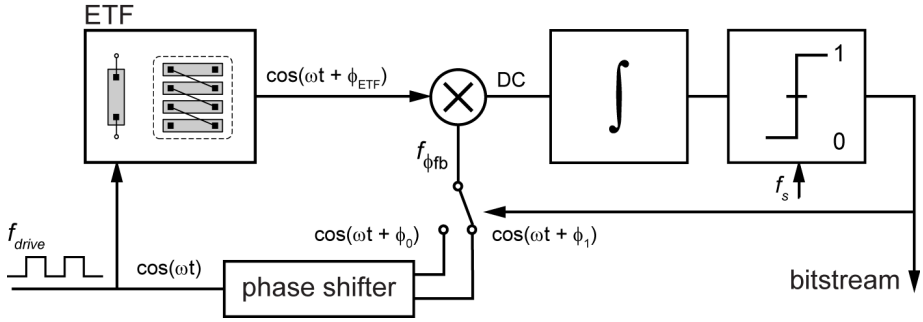


Fig. 3.8: Phase-domain Sigma-Delta Modulator (PDΣΔM) digitizing ϕ_{ETF}.

Assuming that a sampling period is again equal to an integer number of f_{drive} periods, the DC component of the integrator's input as a function of ϕ_{ETF} and the feedback phase shift, ϕ_{fb} (either ϕ_0 or ϕ_1), can be found by integrating over one f_{drive} period:

$$DC = \int_0^{1/f_{drive}} \cos(2\pi f_{drive} t + \phi_{fb}) \cos(2\pi f_{drive} t + \phi_{ETF}) dt = \frac{1}{2} \cos(\phi_{ETF} - \phi_{fb}) \tag{3.3}$$

The PDΣΔM will interpolate between these two DC components. Phase references ϕ_0 and ϕ_1 should be selected such that $\cos(\phi_{ETF} - \phi_0) > 0$, and $\cos(\phi_{ETF} - \phi_1) < 0$. The bitstream average, μ, has a range from 0 to 1 and is given by:

$$\mu = \frac{\cos(\phi_{ETF} - \phi_0)}{\cos(\phi_{ETF} - \phi_0) - \cos(\phi_{ETF} - \phi_1)} \tag{3.4}$$

3.3.3.2 Design Considerations

The design of a typical PDΣΔM for ETF readout is discussed below, using an s=24µm D_{Si} ETF in a 0.18µm CMOS process as an example. From section 2.5.3, such an ETF will achieve good energy efficiency when f_{drive} = 42kHz. At this drive frequency, simulations show that ϕ_{ETF} will vary from about 36º to 70º over the -55ºC to 125ºC range (Fig. 3.9). Choosing e.g. ϕ_1 and ϕ_0 –67.5° and –11.25°, respectively, defines an appropriate phase input range. The f_{ref} required to generate phase shifts at this spacing should be at least f_{drive} *(360/11.25) = 1.365MHz.

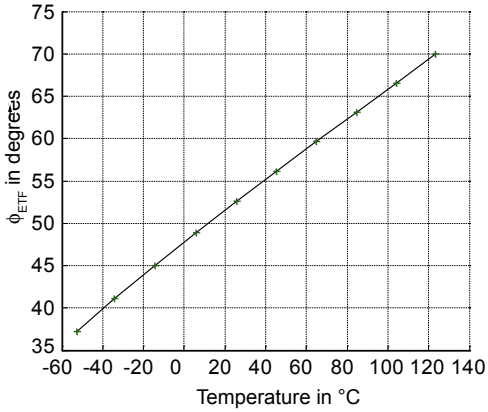


Fig. 3.9: Simulated ϕ_{ETF} for a D_{Si} ETF with s=24µm driven at f_{drive} = 42kHz.

Having set the input range of the PDΣΔM to meet the expected ϕ_{ETF}, other specifications can be defined. Based on the lithographic spread of 0.18µm CMOS technology, a ±0.2ºC (3σ) lithographic spread in *s* is to be expected. This means that all PDΣΔM errors should be at the ±0.02ºC level. Linearizing the Fig. 3.9 curve yields a sensitivity of about 5.5ºC / ºϕ, so all phase errors should be at the 3.7mº level (corresponding to 0.24ns for f_{drive} = 42kHz).

When this ETF is driven at its typical heater power of P_{heat} = 2.5mW, its output amplitude is about $400\mu V_{pp}$. Simulations show that in this case, the sensitivity of the demodulated output is about 350nV/ºC. Assuming that the ETF's resistive thermopile is the dominant source of noise, at $v_{n,Rtp}$ = 18nV/√Hz, a $0.02ºC_{rms}$ resolution requires the bandwidth to be limited to 0.16Hz, or a conversion rate of 0.32S/s. For the 1st-order PDΣΔM to have a 3.7mº phase quantization error after being decimated to this rate by a sinc filter, its OSR needs to be at least 30600; For simplicity, the OSR is set to 32768, which then requires $f_s \geq f_{drive}/4$ = 10.67kHz.

3.3.3.3 Multi-bit PDΣΔM architectures

Although the 1-bit PDΣΔM typically meets all specifications, its OSR has to be rather high to reduce its quantization error to 0.02ºC. Using multi-bit phase feedback can reduce the required OSR.

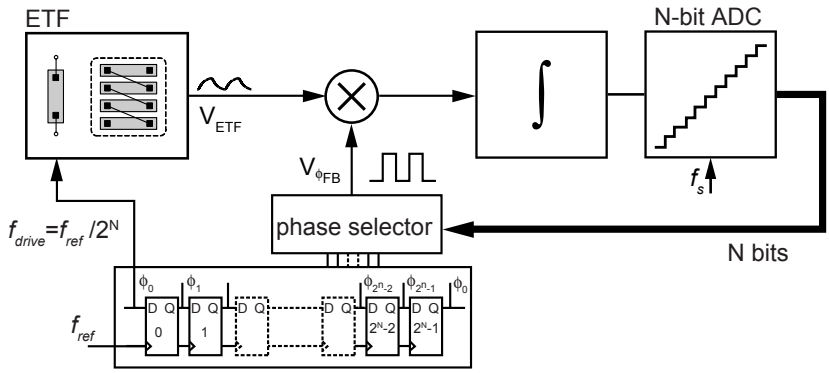


Fig. 3.10: Multi-bit PDΣΔM using a cyclical counter as phase DAC.

Fig. 3.10 shows a multi-bit PDΣΔM. Instead of using a programmable frequency divider (for multi-bit frequency readout), a cyclical counter can be used as a phase DAC. The implementation of such a phase DAC is very straightforward, and the taps of this counter generate linearly spaced phase references. Incrementing N doubles the phase resolution, and thus halves the required OSR. For fixed f_{drive}, this requires a proportionally higher f_{ref}.

The accuracy of the phase DAC will typically be very high. Since a typical thermal delay is much larger than the propagation delay of a flip-flop (~1µs vs. ~50ps in modern CMOS), the phase DAC typically meets the required accuracy; if needed, further reduction of static timing errors can be achieved e.g. by rotating the ϕ_0 starting point through the DFF chain. Clock jitter on the phase references translates into phase noise, but this is averaged by the decimation filter; in most cases, the thermal noise from the ETF is dominant.

The Fig. 3.10 multi-bit PDΣΔM also requires an N-bit ADC. Especially for high N, this is not straightforward to implement. Fortunately, the low bandwidth of most temperature signals can again be leveraged to alleviate this issue, by using a two-step or *zoom* phase-to-digital converter (ZPDC). In a ZPDC, ϕ_{ETF} is first coarsely quantized by a fast ADC using e.g. a SAR or a single slope algorithm to approximate ϕ_{ETF}. Fig. 3.11 shows an example.

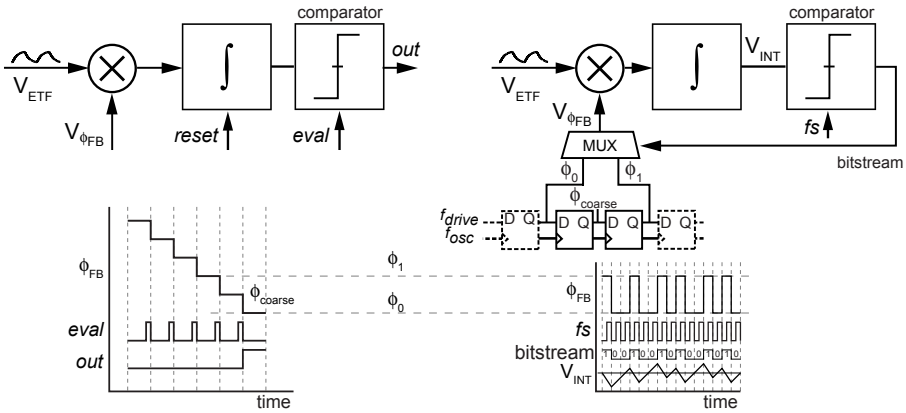


Fig. 3.11: Two-step phase-to-digital converter

Here, ϕ_{ETF} is first coarsely determined by a single-slope ADC and then accurately digitized using a phase-domain sigma-delta modulator (PDΣΔM). During the first step, V_{ETF} is multiplied with another signal at f_{drive} whose phase shift, ϕ_{ramp}, is ramped downwards in ϕ_{step} decrements. For each step, the multiplier output is integrated for several sampling periods, after which a sampled comparator detects the polarity of V_{int}. Since the DC component of the multiplier is proportional to $\cos(\phi_{ETF} - \phi_{ramp})$, the comparator detects a zero crossing when $\phi_{ETF} - \phi_{ramp} \geq 90°$. At this point, the instantaneous value of ϕ_{ramp} is stored as ϕ_{coarse}, and the integrator is reset. In the second digitization step, two phase references ϕ_0, $\phi_1 = \phi_{coarse} \pm \phi_{step}$ define the input range of a 1st-order, single-bit PDΣΔM, whose decimated output ϕ_{fine} represents a weighted average of ϕ_0 and ϕ_1. Since ϕ_{ETF} is already coarsely known, the PDΣΔM's input range can be significantly reduced, to only $\pm 360° * f_{drive}/f_{ref}$, reducing quantization error significantly.

3.3.3.4 Conclusion

This section has provided an overview of some phase digitizer concepts for ETFs. Chapter 4 will discuss how to implement these in CMOS technology. For now, Table 3.2 provides an overview of the specifications for phase-domain ETF readout circuits.

Table 3.2: Top-level specifications for D_{Si} ETF and PDΣΔM / ZPDC

	From ETF (s=24µm)	PDΣΔM	ZPDC
Temp. inaccuracy	±0.2ºC (3σ)	±0.02ºC	
Phase inaccuracy	37mº	3.7mº	
f_{drive}	42kHz is close to optimum (Ch.2)	Sufficient bandwidth to have minimal phase shift at this frequency → ≈ 100MHz	
f_{ref}	≥ f_{drive}	Defined by ϕ_0 and ϕ_1	≥ f_{drive} * 360º/ϕ_{step}
Thermal noise	18nV/√Hz from R_{tp}	9nV/√Hz (half the ETF's noise power)	
Quantization noise	n.a.	Quantization noise: 3.7mº → OSR = 2^{15}	OSR reduced by N in 'zoom-in' step
Linearity	$T^{0.9}$	Ideally, compensates for $T^{0.9}$ D_{Si} ETF nonlinearity	
Temperature range	Arbitrary	-55ºC to 125ºC (application-driven)	
Power consumption	P_{heat} = 2.5mW	$P_{circuit} \leq P_{heat}$	
V_{DD}	Arbitrary, V_{DD} defines R_{heater}	1.8V	
Technology	Arbitrary, set by accuracy target	0.18µm CMOS	

3.4 Frequency references based on a temperature-compensated ETF

ETFs can also be used to realize fully integrated frequency references. Although only tangentially related to this work, the main concept is described below. For further information, the reader is referred to [3.7], in which a much more detailed analysis is presented.

A D_{Si} ETF-based frequency reference employs an FLL similar to the one discussed in section 3.2, including an ADC and a DAC in the filter path. However, rather than keeping ϕ_{ETF} constant (at 90º), a temperature sensor is used to measure the die temperature close to the ETF, to tune the phase set point (ϕ_{SET}) so as to keep f_{osc} constant. In a way, this is an inversion of the phase-domain readout discussed in the previous section: instead of applying a constant f_{drive} and measuring $\phi_{ETF}(T)$, applying the $T^{0.9}$ proportionality of ϕ_{ETF} to ϕ_{SET} leads to a constant output frequency.

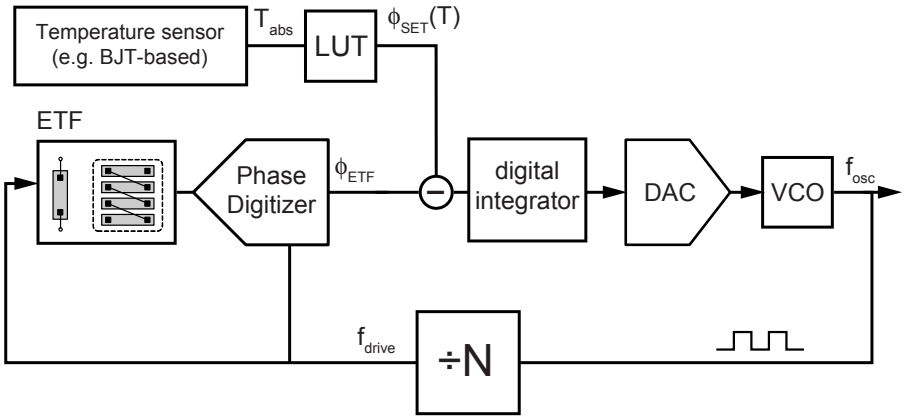


Fig. 3.12: ETF-based frequency reference. An absolute temperature sensor is used to compute a phase shift that compensates for the temperature dependence of ϕ_{ETF}.

$\phi_{SET}(T)$ is applied to the loop by mapping the temperature sensor's output onto a polynomial, after which the resulting code is subtracted from the input of the digital integrator (see Fig. 3.12). The temperature sensor should have a low inaccuracy and be in close proximity to the ETF. BJT-based sensors are well-suited for this, since their accuracy can be as low as ±0.1ºC (3σ) [3.9], and since they can be integrated on the same die, right next to the ETF. BJT-based sensors can interface with a ΣΔM to output a digital representation of absolute temperature. This allows the mapping of T_{abs} to ϕ_{SET}(T) to be done in the digital domain, so that high accuracy can be achieved.

The accuracy requirements on this ETF system are similar to those for a DAFLL or a PDΣΔM. The DAC needs to have sufficient range to compensate for the process, voltage and temperature dependence of the oscillator, and the resolution of the digital integrator and the DAC need to be sufficiently high to achieve small steady-state error and cycle-to-cycle jitter.

Noise in this system manifests itself as jitter at the VCO output. At low frequencies, where the loop gain is high, the ETF dominates the noise performance, while at high frequencies, the noise of the DAC and oscillator are more significant.

A state of the art D_{Si} ETF-based frequency reference [3.7] has an output frequency f_{osc} of 16MHz and achieves an inaccuracy of ±0.1% over the military temperature range (-55°C to 125°C). Its BJT-based temperature sensor had a trimmed inaccuracy of ±0.2ºC (3σ), which contributed about half of its frequency inaccuracy. For a power consumption of 2.1mV, its jitter is $45ps_{rms}$ (at room temperature).

3.5 Temperature-to-digital converters based on both D_{Si} and D_{SiO2} ETFs

3.5.1 General concept

In many applications, e.g. the thermal management of µP's, an accurate time reference is available, and so temperature sensors based on a single *silicon* ETF can be realized. However, in stand-alone applications, such a reference may not be available. Commercially available stand-alone temperature-to-digital converters rely on the ratiometric measurement of V_{PTAT}/V_{BG}, which are both derived from on-chip BJTs and so do not require an external reference. In order to be of interest for stand-alone applications, ETF-based temperature sensors should also be self-referenced.

Conceptually, a self-referenced ETF-based temperature sensor should output a digital signal represented by the following equation (phase-domain readout is assumed):

$$T_{abs} = \alpha(T)\frac{\phi_{ETF}(T)}{\phi_{REF}} \tag{3.5}$$

Here, ϕ_{ETF}(T) can be the temperature-dependent phase shift of a *silicon* ETF, while $\alpha(T)$ is a non-linear gain term that corrects for the $T^{0.9}$ nonlinearity of ϕ_{ETF}, and ϕ_{REF} is a phase reference that is generated on-chip, e.g. by an oscillator that generates f_{drive} ($\phi_{REF} \propto 1/f_{drive}$).

From Eq. (3.1), it can be shown that a temperature error of ±0.05ºC requires f_{ref} to have a maximum absolute error of ±225ppm (at worst case, T=125ºC).

Silicon ETFs are strongly temperature-, but also frequency-dependent. As mentioned before, a ±225ppm frequency error leads to a ±0.05ºC temperature error, or, perhaps more strikingly, achieving errors below ±1ºC requires the inaccuracy of f_{drive} to be better than ±0.5% over PVT. In standard CMOS technology, it is quite challenging to implement such an accurate oscillator, and this typically requires the use of complex trimming and temperature compensation techniques ([3.13][3.14] and the overview in [3.7]). Such oscillators also consume significant chip area and power, further reducing the attractiveness of such a system.

A more elegant solution is to implement the time reference in the thermal domain. In Chapter 2, it was noted that D_{SiO2} is about 20x less temperature dependent than D_{Si}, making $\phi_{ETF,SiO2}$ a good candidate. However, it was also noted that measuring it typically requires a more complicated set-up involving the differential measurement of two ETFs. One option is to use an ETF with both a polysilicon and a diffusion heater, the former thermal path being longer by the field oxide / STI. In SOI technology, another option is to place a deep trench isolation structure in the thermal path, since the trench is (partially) made of SiO_2.

Self-referenced sensors will thus typically consist of two ETFs. The first, ETF1, will have a thermal path consisting of only silicon, and so ϕ_{ETF} can be approximated by:

$$\phi_{ETF1} \propto s_1\sqrt{f_{drive}/D_{Si}}\ , \tag{3.6}$$

while ETF2 will have the same plus a thermal path in SiO_2, resulting in:

$$\phi_{ETF2} \propto s_2\sqrt{f_{drive}/D_{Si}} + t_{ox}\sqrt{f_{drive}/D_{SiO2}} \tag{3.7}$$

One implementation of a sensor combining these two ETFs is shown in Fig. 3.13:

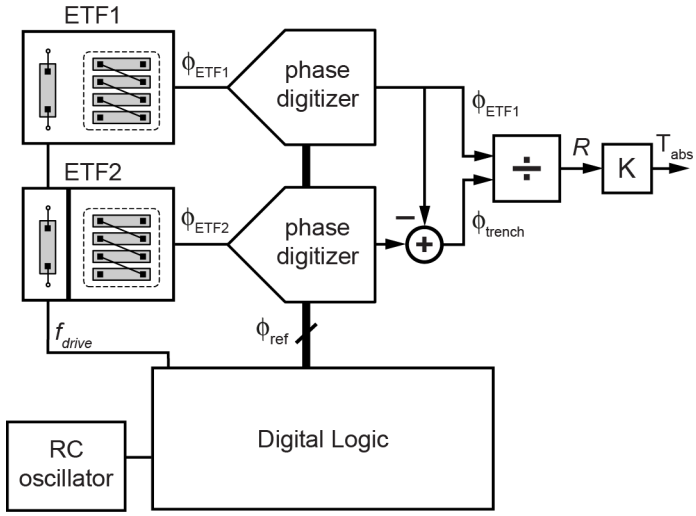


Fig. 3.13: Ratiometric temperature sensor based on D_{Si} and D_{SiO2}.

Here, both ETFs are driven by the same f_{drive}, and their phase shifts are digitized using phase digitizers (e.g. PDΣΔMs). T_{abs} is of the form:

$$T_{abs} = K \frac{\phi_{ETF1}}{\phi_{ETF2} - \phi_{ETF1}} \tag{3.8}$$

A closer look at ratio R yields:

$$R = \frac{\phi_{ETF1}}{\phi_{ETF2} - \phi_{ETF1}} = \frac{s_1 \sqrt{f_{drive} / D_{Si}}}{s_2 \sqrt{f_{drive} / D_{Si}} + t_{ox} \sqrt{f_{drive} / D_{SiO2}} - s_1 \sqrt{f_{drive} / D_{Si}}} , \tag{3.9}$$

Which, assuming $s_1 = s_2 = s$ for now, reduces to:

$$R = \frac{s}{t_{ox}} \sqrt{\frac{D_{SiO_2}}{D_{Si}}} \propto T^m \tag{3.10}$$

As the above equation shows, this ratio is only a function of geometry and the ratio of thermal diffusivities (and not of f_{drive}). Neglecting thermal expansion, the exponent m is the difference between the exponents of the D_{Si} and D_{SiO2} T^n power laws; since D_{SiO2} is nearly constant, $m \approx 0.9$.

Consider two s=8µm ETFs in SOI CMOS, one having a deep isolation trench with an effective SiO_2 path of about 1µm in the heat diffusion path. Fig. 3.14 shows the simulated phase shifts:

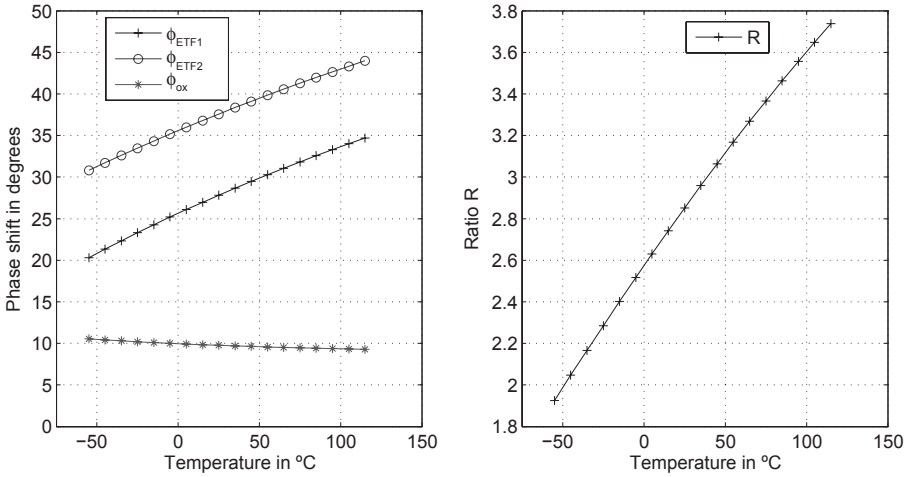


Fig. 3.14: Simulated phase shifts and their ratio as a function of temperature.

As the above figure and Eq. (3.10) show, R, like ϕ_{ETF}, is also a near-linear function of temperature that should have a reduced sensitivity to variations in f_{drive}.

3.5.2 f_{drive} sensitivity

The above equations show that the sensitivity to f_{drive} is cancelled to first order, so that an inaccurate RC oscillator can be used to generate f_{drive}. In practice, second order effects such as the presence of higher harmonics in the heater drive signal and geometrical differences in the ETF's thermal path lead to deviations from Eqs. (3.6), (3.7), and thus to finite f_{drive} cancellation. Section 5.4.2 will present the measured results for a testchip; at about 100x, the reduction in f_{drive} sensitivity reduces the requirements on the oscillator.

Even though the requirements on the on-chip oscillator are reduced, finite f_{drive} cancellation still sets some specifications for the oscillator. For a ±0.05ºC error, the oscillator should have an error of no more than ±2%, thus still requiring a reasonably accurate implementation. The oscillator will have both random (device-to-device) spread and a temperature coefficient. While the nominal temperature coefficient does not introduce spread, it will affect curvature in R, and therefore linearity. It is also a potential source of wafer-to-wafer or lot-to-lot spread. Random spread causes an unmitigated error that can be reduced by trimming, if required.

3.5.3 Linearity

The linearity of R is determined by Eq. (3.9), which contains D_{Si}, D_{SiO2} and geometry components. In the simple case of $s_1 = s_2$ and $dD_{SiO2}/dT = 0$, R will have the nonlinear $T^{0.9}$ D_{Si} temperature dependence. Interestingly, the additional degree of freedom in choosing $s_1 \neq s_2$ can be used to adjust the curvature of R so as to make it more linear. Fig. 3.15 shows the linear error in R as a function of slightly increasing s_1 for fixed s_2. It can be seen that optimal linearity is achieved for $s_1 = 8.55\mu m$. In practice, second order effects not included in Eq. (3.9) also affect R, so that the optimal s_1 is best found by characterization (see Ch. 5).

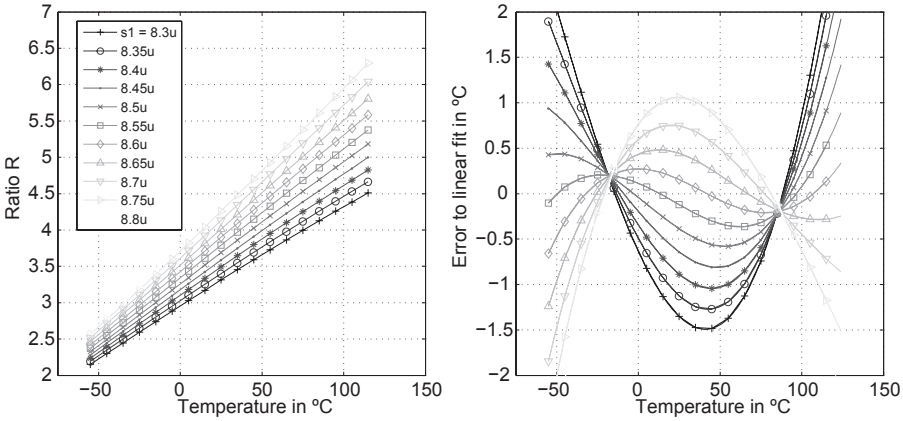


Fig. 3.15: Simulated ratio *R* and linear error (in ºC) as a function of s_1.

In general, s_1 should always be selected to optimize linearity; the slight noise penalty in going from s=8.0 to 8.55μm is typically less serious a problem than that of (digitally) linearizing a $T^{0.9}$ temperature dependence.

A potential source of error is the nonlinearity of the phase digitizers. Firstly, this may affect R; secondly, for the f_{drive} cancellation suggested by Eq. (3.9) to work, the phase digitizers should not distort the phase shifts that they measure. In practice, the linear error of the phase digitizers should be comparable to their resolution (see Chapter 4 for more details).

3.5.4 Accuracy

In section 3.4, the key specifications for the phase digitizer (PDΣΔM) were derived for a single D_{Si} ETF temperature sensor. These specifications will now be revisited for the two-ETF system. Intuitively, since determining R requires the combination of phase measurements on two ETFs, more stringent requirements on their phase error can be expected. From Eq. 3.8, the temperature error due to deterministic errors in ϕ_{ETF1} and ϕ_{ETF2} is given by:

$$dT = K\frac{\phi_{ETF1}\phi_{ETF2}}{\phi_{ETF2}-\phi_{ETF1}}\left(\frac{d\phi_{ETF1}}{\phi_{ETF1}}-\frac{d\phi_{ETF2}}{\phi_{ETF2}}\right) \tag{3.11}$$

And for random errors (e.g. noise, lithographic spread), by:

$$dT = K\frac{\phi_{ETF1}\phi_{ETF2}}{\phi_{ETF2}-\phi_{ETF1}}\sqrt{\left(\frac{d\phi_{ETF1}}{\phi_{ETF1}}\right)^2+\left(\frac{d\phi_{ETF2}}{\phi_{ETF2}}\right)^2} \tag{3.12}$$

Eq. (3.11) can be split into two and, defining $\phi_{Si} = \phi_{ETF1}$ and $\phi_{ox} = (\phi_{ETF1} - \phi_{ETF2})$, approximated by:

$$dT = K\frac{\phi_{Si}}{\phi_{ox}}d\phi_{ETF1} + K\frac{\phi_{Si}}{\phi_{ox}}d\phi_{ETF2} \tag{3.13}$$

This shows that a constant absolute phase error (e.g. $d\phi_{ETF1}$) leads to a *temperature-dependent* temperature error, since ϕ_{ox} is nearly constant while ϕ_{Si} has a $T^{0.9}$ temperature dependence. Clearly, both ϕ_{Si} and ϕ_{ox} errors are most significant at high temperature.

A numerical example is used to illustrate the implications for the specifications for the ETFs and their phase digitizers. Fig. 3.16 shows the effect of a constant 10mº phase error for both ϕ_{Si} and ϕ_{ox}.

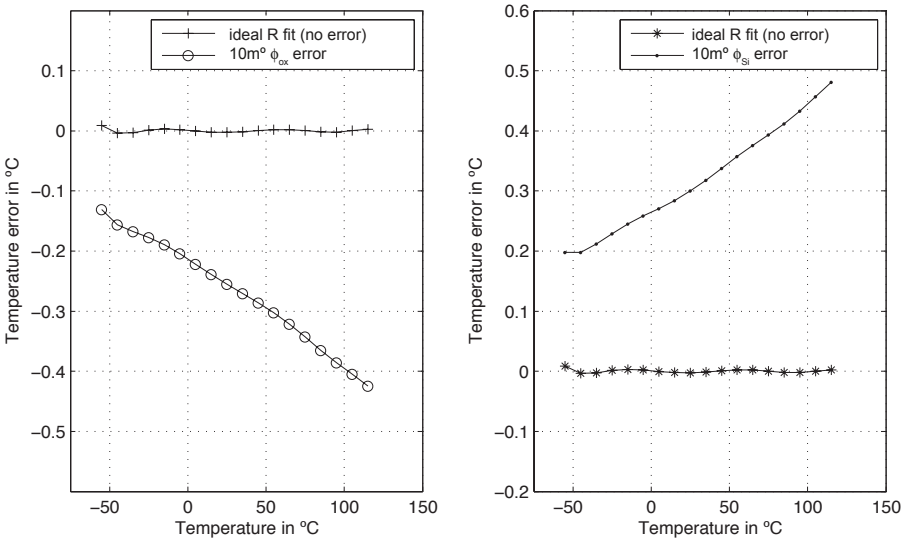


Fig. 3.16: Simulated temperature error for phase errors in ϕ_{Si} and ϕ_{ox}.

The above figure shows a worst-case error of up to 0.5ºC. Comparatively, a linear fit to ϕ_{ETF1} in Fig. 3.14 reveals a temperature sensitivity of about 11ºC/ºϕ, so the same 10mº phase error there only corresponds to an error of about 0.1ºC. Through its ϕ_{Si}/ϕ_{ox} ratio, the self-referenced sensor is more sensitive to phase errors than a comparable single silicon ETF system, in this example by about 5 times. Consequentially, errors in the phase digitizers need to be proportionally smaller to achieve the same temperature-sensing inaccuracy.

The above also shows that having a small ϕ_{ox} compared to ϕ_{si} undesirably amplifies the errors from both ϕ_{ETF}'s; setting $\phi_{si} \approx \phi_{ox}$ is preferred. Through the ratio of their diffusivities, this implies that $s \approx 10t_{ox}$. For a typical s as used in the previous examples, t_{ox} must be increased (e.g. by diffusing heat through multiple oxide layers), but this likely decreases the amplitude of the signal used to detect ϕ_{ETF2}, increasing noise. A better way is to scale s, although lithographic spread and process design rules do not always allow for this.

While this tradeoff can potentially be solved by technology improvement, a more serious source of error is spread in ϕ_{ox}. While the absolute tolerance in t_{ox} is typically very small (i.e. at the nanometer level), the relative error in ϕ_{ox} for a thin oxide layer can be quite significant (i.e. several %). From the above, a 1% error in ϕ_{ox} causes an error of about ±5ºC, thus overwhelming any ϕ_{Si} or PDΣΔM phase error.
An effective way to reduce all these errors is to apply a temperature-dependent trim to R. From Fig. 3.16, errors in ϕ_{Si} and ϕ_{ox} both are strongly temperature-dependent, crossing zero error at about -125ºC, or T_0. An effective gain trim to R is of the form:

$$R_{trimmed} = R + \alpha(T - T_0) \tag{3.14}$$

The value of T_0 is determined by geometry and thermal diffusivities, and will therefore be relatively constant between devices, so that only a single-temperature trim is required. Measurements in Chapter 5 will show that applying this trim enables the realization of high-precision temperature sensors over a wide range.

Having presented both the single D_{Si} ETF and the self-referenced D_{Si} and D_{SiO2} ETF sensors above, the need for precision phase digitizers such as the PDΣΔM and the ZPDC has become apparent. The next chapter will discuss how to implement such phase digitizers in standard CMOS technology.

3.6 References

[3.1] V. Szekely, “Thermal monitoring of microelectronic structures”, *Microelectronics J.*, vol. 25, no. 3, pp. 157–170, 1994.

[3.2] V. Szekely and M. Rencz, “A new monolithic temperature sensor: the thermal feedback oscillator,” *Digest Transducers*, Jun. 1995, vol. 15, pp. 849–852.

[3.3] A. Abidi, R.G. Meyer, “Noise in relaxation oscillators”, *IEEE J. Solid-State Circuits*, vol. 18, no. 6, pp. 794-802, Dec. 1983.

[3.4] K.A.A. Makinwa and J.F. Witte, “A temperature sensor based on a thermal oscillator,” *Proc. IEEE Sensors*, pp. 1149–1151, Oct. 2005.

[3.5] K.A.A. Makinwa and M.F. Snoeij, “A CMOS temperature-to-frequency converter with an inaccuracy of ±0.5°C (3σ) from -40 to 105°C,” *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2992–2997, December 2006.

[3.6] C. Zhang and K.A.A. Makinwa, “Interface Electronics for a CMOS Electrothermal Frequency-Locked-Loop”, *IEEE J. Solid-State Circuits*, vol. 43, no. 7, pp. 1603–1608, July 2008.

[3.7] S. M. Kashmiri, “Thermal-Diffusivity-Based Frequency References in Standard CMOS”, Ph.D. Thesis, Delft University of Technology, Apr. 2012.

[3.8] S. M. Kashmiri, “A digitally-assisted electrothermal frequency-locked loop”, *Proc. ESSCIRC 2009*, pp. 296-299, Sept. 2009.

[3.9] M.A.P. Pertijs, K.A.A. Makinwa and J.H. Huijsing, “A CMOS temperature sensor with a 3σ inaccuracy of ±0.1ºC from -55ºC to 125ºC”, *IEEE J. Solid-State Circuits*, pp. 2805-2815, Dec. 2005.

[3.10] I. Galton and G. Zimmerman, “Combined RF phase extraction and digitization”, *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 1104–1107, May 1993

[3.11] I. Galton et al., “A Delta-Sigma PLL for 14-b, 50kSample/s Frequency-to-Digital Conversion of a 10MHz FM Signal”, *IEEE J. Solid-State Circuits*, Vol. 33, No. 12, Dec. 1998

[3.12] K.A.A. Makinwa, M.A.P. Pertijs, J.C. van der Meer and J.H. Huijsing, “Smart sensor design: the art of compensation and cancellation”, *Proc. ESSCIRC*, pp. 76 - 82, Sept 2007.

[3.13] M. S. McCorquodale et al., “A silicon die as a frequency source,” in Proc. IEEE Int. Frequency Control Symp., 2010, pp. 103–108.

[3.14] A. Ahmed et al., "A highly stable CMOS Self-Compensated Oscillator (SCO) based on an LC tank temperature Null concept," *Proc. IEEE IFCS-EFTF,* pp. 1-5, May 2011.

4 A Phase-Domain Sigma-Delta Modulator for ETF readout

4.1 Introduction

As stated in Chapter 1, ETFs will only become competitive temperature sensors when they are combined with suitable readout circuits. The previous chapter has shown that for both crystal-referenced and self-referenced ETF systems, such circuits should be high resolution, narrowband frequency- or phase-to-digital converters. It was also shown that the phase-domain sigma-delta modulator (PDΣΔM) is the best candidate for ETF readout, since operating the ETF at a constant f_{drive} achieves the highest linearity.

Phase digitizers for ETFs can be implemented according to various application requirements. In µP thermal management, it is advantageous to optimize for circuit area and sampling rate, possibly at the cost of precision and resolution. For ETF characterization and model development, the requirements are different: the phase digitizer should be 'transparent', so that phase errors can be reliably attributed to the ETF. In this case, circuit area, power consumption and conversion rate are less important.

This chapter will follow the latter approach and discuss the design of PDΣΔMs for ETF characterization. Circuit techniques used to achieve high precision and high resolution will be presented, using an implementation in a 0.18µm CMOS process as an example. Section 4.2 will discuss the implementation of the PDΣΔM, starting with a system-level approach and then discussing each block in greater detail. Section 4.3 will focus specifically on how to achieve high precision using dynamic error cancellation techniques. Section 4.4 will analyze the PDΣΔM's cosine nonlinearity. The zoom-in phase-to-digital converter (ZPDC) removes this nonlinearity while offering multi-bit quantization; this is presented in section 4.5. Section 4.6 discusses several ways to linearize the temperature sensor's output. Section 4.7 concludes the chapter with measurements on a stand-alone PDΣΔM, confirming that the circuit meets the requirements.

4.2 PDΣΔM Implementation

The example PDΣΔM discussed in this chapter will be designed to interface an s=24µm ETF in a 0.18µm CMOS process. A similar PDΣΔM will be needed to test ETFs in other technologies (such as 0.7µm Bulk and 0.5µm SOI CMOS), but the ETFs in 0.18µm CMOS have the lowest expected inaccuracy due to lithographic spread, and thus pose the biggest challenge to PDΣΔM design. The numbers in the following section are based on circuits and simulations in 0.18µm CMOS; similar designs in the other technologies have achieved comparable results [4.10][4.11].

4.2.1 Overview

The previous chapter has already presented several phase digitizers and some of their associated system-level considerations. The block diagram of a 1st-order, single-bit PDΣΔM is repeated below:

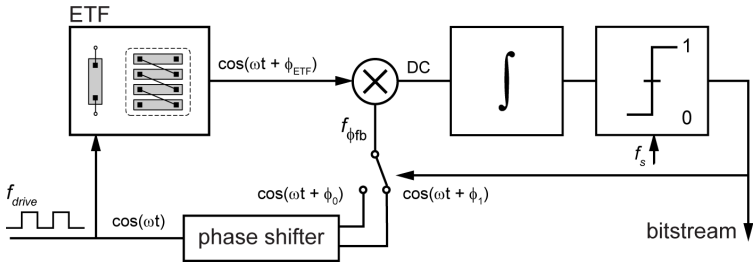


Fig. 4.1: Phase-domain Sigma-Delta Modulator (PDΣΔM) digitizing ϕ_{ETF}.

The need for negligible error from the PDΣΔM translates into demanding specifications, because the ETF's expected inaccuracy is only ±0.15°C (3σ). At f_{drive} = 85kHz, ϕ_{ETF} changes by about 36 degrees over the –55 to 125ºC range, which translates into a linearized sensitivity of ≈5ºC/ºϕ. For ±0.015ºC error, PDΣΔM's phase errors should be at the 0.003ºϕ level, and its resolution should be greater than 13.5 bits. Table 4.1 is a more extended version of Table 3.2 and shows these specifications, for both the PDΣΔM and the ZPDC.

Table 4.1: Overview of PDΣΔM specifications:

	Conditions for an *s*=24μm Si ETF	PDΣΔM	ZPDC
Temp. inaccuracy	±0.15ºC (3σ)	±0.015ºC	
Phase inaccuracy	3mº	3mº	
f_{drive}	85.33kHz is close to optimum (Ch.2)	Sufficient bandwidth to have little phase shift at this frequency → ≈ 100MHz	
Timing error at f_{drive}	For 3mº phase error	**0.13ns**	
f_{ref}	≥ f_{drive}	Defined by ϕ_0 and ϕ_1	≥ f_{drive} * 360º/ϕ_{step}
Thermal noise	18nV/√Hz from R_{tp}	9nV/√Hz (a quarter of the ETF's noise power)	
Quantization noise	n.a.	Quantization noise: 3.7mº → OSR = 2^{15}	OSR reduced by N in 'zoom-in' step
ΣΔM sampling frequency (f_s) for quantization noise ≈ thermal noise	n.a.	1st-order PDΣΔM: 10kHz	625Hz (for N = 16)
Offset at demodulator output	At P_{heat} = 2.5mW	**I_{OS} < 7pA**	
Offset referred to V_{ETF}	At P_{heat} = 2.5mW	**V_{OS} < 12nV**	
Linearity	$T^{0.9}$	Ideally, compensates for $T^{0.9}$ D_{Si} ETF nonlinearity ±0.1ºC to ±0.2ºC linear error is acceptable	
Temperature range	Arbitrary	-55ºC to 125ºC (application-driven)	
Power consumption	P_{heat} = 2.5mW	$P_{circuit} \leq P_{heat}$	
V_{DD}	Arbitrary, V_{DD} defines R_{heater}	1.8V	
Technology	Arbitrary, set by accuracy target	0.18μm CMOS	

The remainder of this chapter discusses how to meet these specifications. Because V_{ETF} is very small (at only several 100μV_{pp}), the offset requirements are very challenging: referred to a demodulated V_{ETF}, V_{OS} should be no more than 12nV. Moreover, at 130ps, the timing accuracy needs to be rather high, translating to high-bandwidth processing of V_{ETF}. Taking the above into consideration leads to a

precision readout circuit with a high-bandwidth input stage. A block-level implementation is shown in Fig. 4.2:

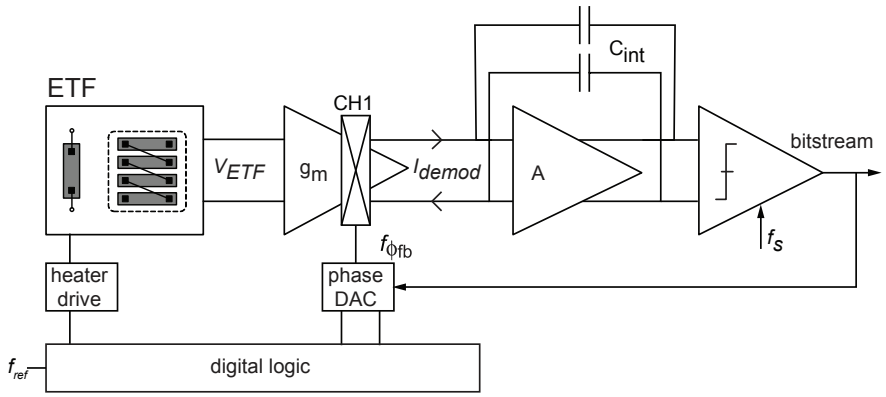


Fig. 4.2: Block-level diagram of a PDΣΔM reading out an ETF.

Following the signal path from left to right:

- The ETF is driven by a synchronized heater drive circuit
- The ETF output, V_{ETF}, is converted to a current by a wide-bandwidth g_m-stage
- The current is demodulated by a chopper (CH1), driven by the phase DAC
- The demodulated current (I_{demod}) is integrated using an active integrator
- A comparator implements the PDΣΔM's 1-bit ADC

Digital logic divides down f_{ref} to create the f_{drive} and feedback phase shifts signals. The bitstream is decimated by a sinc filter (not shown) to obtain a high-resolution digital output. The modulator can be run continuously or incrementally, depending on the system requirements.

The Fig. 4.2 topology was chosen for its power efficiency, its low voltage headroom and its potential for accuracy. A g_m-stage drives C_{int} with a current proportional to the demodulated V_{ETF}. The g_m-stage offers low noise and high bandwidth, while still being sufficiently linear because V_{ETF} is small. Its high input impedance ensures that the ETF's thermopile is only capacitively loaded, while a high output impedance helps to increase the DC gain of the integrator.

The use of an active integrator decouples the signal swing from the g_m-stage's outputs and increases the overall DC gain. This simplifies the design of the g_m-stage, which now drives a virtual ground so that it can use large overdrive voltages for high gain and improved matching. The second-stage opamp, on the other hand, does not require low offset or noise, but can have a rail-to-rail output swing to maximize the available signal swing, making this topology suitable for scaled technologies that have low $V_{DD,max}$.

The value for g_m is determined by noise, while the value for C_{int} will be determined by the allowable signal swing and by f_s (since this is a continuous-time ΣΔM). The swing is determined by integrating the PDΣΔM's error signal, and has a peak-to-peak amplitude of approximately:

$$V_{swing,pk-pk} = A\frac{8}{\pi^2}\cos\left(\phi_1 - \phi_0\right)\frac{g_m}{C_{\text{int}} f_s}, \tag{4.1}$$

in which A is the peak-to-peak amplitude of V_{ETF}. C_{int} was chosen to be 25pF; while several times larger than required, this provides flexibility during test, since it ensures that the ΣΔM does not clip

when reading out an ETF at high P_{heat} / low f_s. For 200μV_{pp} ETF output sampled at f_s=10kHz, $V_{swing,pk\text{-}pk}$ is about 400mV differentially.

The total DC gain is very high, so that the effect of integrator leakage is sufficiently small. This is important, since integrator leakage causes deadbands in the transfer function of a first-order ΣΔM. Expanding upon [4.9], for the Fig. 4.2 circuit, the width of these deadbands is given by:

$$\Delta x = \frac{1-p}{1+p}\text{, with } p = \exp\left(\frac{-R_{out,ota}C_{\text{int}}}{f_s A_{DC,opamp}}\right) \tag{4.2}$$

For typical values of $R_{out,ota}$ (1e9Ω), C_{int} (10pF), $A_{DC,opamp}$ (100dB) and f_s (10kHz), these are at the 20-bit level, and thus negligible.

Other advantages of the fully differential architecture are rejection of common-mode and power supply noise. The topology is well-suited to offset-cancellation techniques, as will be shown in Section 4.3.2. First, each of the signal path functions in Fig. 4.2 will be discussed individually.

4.2.2 Heater drive

The first step in interfacing an ETF is driving its heater with a signal having a well-defined frequency f_{drive} and phase shift ϕ_{drive}. Any error in ϕ_{drive} will appear as ϕ_{ETF}, and so the heater drive circuitry should have a well-defined delay, while still being able to generate several mW's of heat at frequencies of around 100kHz.

For ETF characterization, the easiest way is to not implement any heater drive circuit and connect the ETF heater terminals straight to IC pins, where they can be driven by a fast off-chip heater drive circuit. This is schematically shown in Fig. 4.3. This method is flexible, as arbitrary waveforms can be applied to the heater and P_{heat} is easily adjusted. For stand-alone operation, however, this method is not practical. Moreover, spread in the on-chip timing logic will appear as spread in ϕ_{ETF}, thus contributing error.

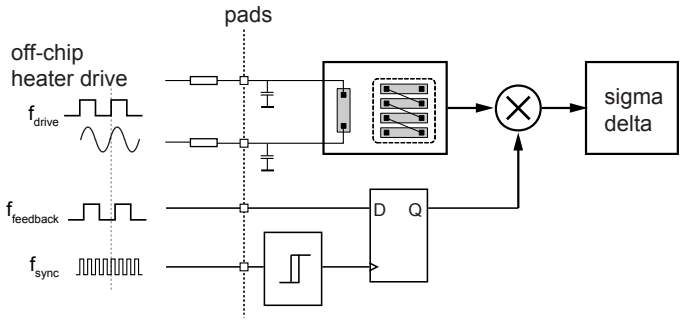
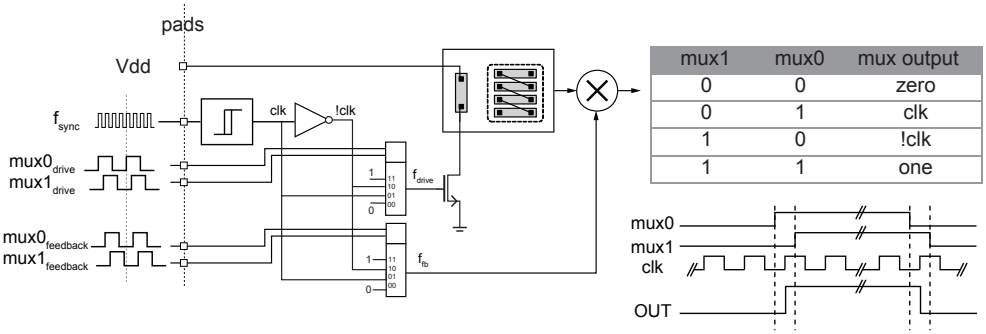


Fig. 4.3: Driving the heater with an off-chip signal.

A simple way of driving ETFs on-chip is by applying a square wave to the heater. This is straightforwardly implemented by placing an NMOS switch in series with the heater. Its gate can be driven by a digital signal, so that no complex driver circuitry is needed, requiring little power and area overhead. The main disadvantage is that some of the heater power will now be at higher f_{drive} harmonics; these are more strongly filtered by the ETF and thus will not contribute as much to the output signal, thus reducing SNR as compared to the ideal case described in section 2.8.

An elegant way of driving the heater with good timing accuracy is by using an 'edge combiner', which is basically a high-speed multiplexer that connects the well-defined edges of a reference clock

signal to the heater switch as needed. Fig. 4.4 shows a circuit-level implementation together with a timing diagram:



mux1	mux0	mux output
0	0	zero
0	1	clk
1	0	!clk
1	1	one

Fig. 4.4: Edge combiners create f_{drive} and f_{fb} to drive the heater and the demodulator with high timing accuracy.

The same circuit is used for both the heater and the demodulator. Each multiplexer can switch to static high, static low, clk and clk inverted. The multiplexer is switched on the falling edge of f_{sync}, so a settling time of $1/2f_{sync}$ is available to switch to the correct state. Since there is no additional logic in the critical delay chain, excellent timing accuracy can be achieved. This approach requires a high frequency f_{sync}, but this signal is typically already required to generate precision phase shift references.

The residual timing spread is associated with the NMOS switches in the heater and the demodulator path. Simulations show that the device-to-device spread on this is about 30ps (3σ), which causes negligible error for an s=24µm ETF. For scaled ETFs, timing accuracy becomes more critical: ϕ_{ETF} is a linear function of s, and SNR is optimized by increasing f_{drive} (see Section 2.5.3.1); this only increases ϕ_{ETF} by a square root proportionality, making the relative effect of a fixed digital delay more significant.

The NMOS switch represents a second source of heat that is also switching at f_{drive}. Although its R_{on} should be much smaller than R_{heater}, the thermopile will still detect the heat dissipated in the transistor, changing the effective s and thus ϕ_{ETF}. The NMOS switch should be designed so that its R_{on} has low spread, and it should be implemented on one of the ETF's axes of symmetry.

4.2.3 Transconductor and demodulator

4.2.3.1 Circuit implementation

V_{ETF} can be converted to a current by either resistors or active devices. Since V_{ETF} is small, a g_m-stage can be used without affecting linearity, while having the advantage of a high-impedance input that does not load the ETF[3].

[3] In principle, the thermopile resistors themselves can be used to perform the V-to-I conversion by connecting the ETF output to a virtual ground. Since this adds no extra noise sources, such a topology potentially has better SNR. In practice however, the poorly defined input impedance at f_{drive} and potentially large offset currents (V_{OS}/R_{TP}) make this difficult to implement.

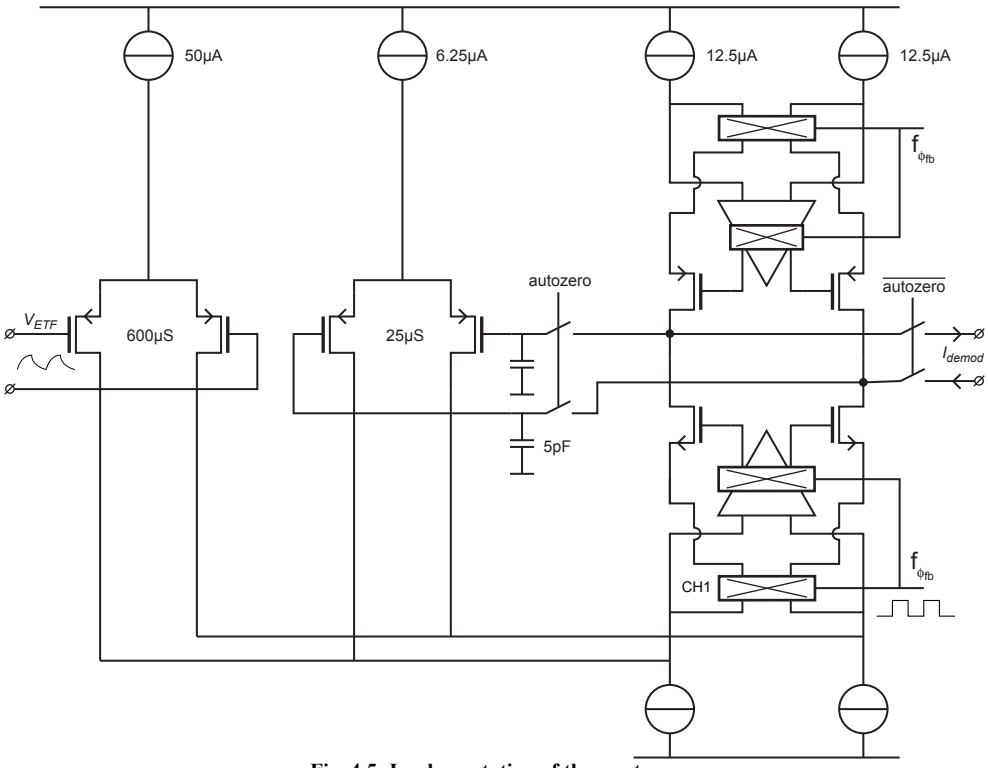


Fig. 4.5: Implementation of the g_m-stage.

Fig. 4.5 shows the schematic of the g_m-stage. It consists of a gain-boosted folded cascode with an auxiliary autozeroing stage [4.4]. The input differential pair converts V_{ETF} to a differential current; this current is demodulated by chopper CH1, which is driven by the phase feedback signal of the PDΣΔM. The demodulated current, I_{demod}, flows into the integrator through the gain-boosted cascodes [4.3]; these ensure that the OTA's output impedance (and thus its DC gain) is sufficiently high.

4.2.3.2 Noise

The value of g_m is mainly determined by noise requirements. Assuming a differential pair in strong inversion and neglecting other noise sources in the circuit, a circuit thermal noise power less than a quarter of the thermal noise power of R_{tp} requires:

$$\overline{v_{n,gm}^2} = \frac{16kT}{3g_m} < \frac{4kTR_{tp}}{2} \Rightarrow g_m > \frac{16}{3R_{tp}} \tag{4.3}$$

For an s=24µm ETF in a 0.18µm process, R_{tp} is about 16kΩ. This translates into a minimum g_m of 330µS. For ETFs with smaller s, R_{tp} will be lower, thus requiring a higher g_m. To also characterize scaled ETFs to compare resolution, g_m was set to 600µS.

Moderate inversion ($g_m/I_D \approx 12$) provides a good tradeoff between input capacitance, linearity and current efficiency, and gives I_D=50µA. A bias circuit adjusts this current over temperature so as to keep g_m constant over temperature [4.7].

4.2.3.3 Autozeroing

The thermal offset in V_{ETF} due to self-heating, the offset and 1/f noise of the input pair, the NMOS and the PMOS current sources all contribute to an error current, i_{error}, at the output of the g_m-stage which can be several μA for worst-case device offset. This current is much larger than the typical signal currents (tens of nA), and would cause a large error in ϕ_{ETF}. Fortunately, the chopper that demodulates V_{ETF} to DC also modulates i_{error} to AC, so that it, in principle, does not introduce DC error. However, the chopped i_{error} still creates a voltage ripple on C_{int}, with amplitude equal to:

$$V_{ripple} = \frac{i_{error}}{2 f_{drive} C_{\text{int}}} \tag{4.4}$$

For typical values of f_{drive} and C_{int} (set by the ETF) and worst-case i_{os}, V_{ripple} can be as much as 1V, in which case it significantly reduces the available signal swing. Moreover, when phase feedback is applied at CH1, the ripple waveform is also phase-shifted; the 1-bit comparator then samples the ripple at a non-zero and feedback-dependent point in time: this adds a small feedforward coefficient to the ΣΔM transfer function, which reduces noise shaping at high frequencies.

The signal swing can be recovered by periodically autozeroing the g_m-stage to reduce i_{error}. The auxiliary auto-zeroing circuitry is also shown in Fig. 4.5. When the autozeroing switches are closed, the OTA operates as a g_m/g_m amplifier that nulls the differential current flowing into its output. Steady state is achieved when the voltage across the auxiliary differential pair, V_{AZ}, is equal to:

$$V_{AZ} = \frac{g_m}{g_{m,az}}(V_{OS} + V_{ETF,DC}) \tag{4.5}$$

Where V_{OS} is the input-referred offset of unchopped g_m-stage. In this circuit, $g_{m,az}$ is 25μS, so V_{AZ} is 24 times larger than (V_{OS}+$V_{ETF,\text{DC}}$). The auxiliary pair is implemented with long devices so that they operate in strong inversion; this maximizes $g_m/g_{m,\text{az}}$ and thus minimizes the input-referred noise and residual offset of the auxiliary differential pair.

During autozeroing, the choppers on the NMOS side should be in their non-crossed state, since they would otherwise periodically invert the polarity of the autozeroing loop. The choppers on the PMOS side should ideally be running as in normal operation, since autozeroing would otherwise overcompensate for the offset current of the PMOS current sources. The turnoff moment of the autozero switches should be timed to correspond to the time at which the ripple induced by the chopped offset of the PMOS current sources is zero.

In contrast to most autozeroing circuits, the OTA inputs are not shorted during autozeroing. This is because the DC component of V_{ETF} ($V_{ETF,DC}$) is a significant source of offset (and therefore ripple), which should also be reduced. The autozeroing loop does not distinguish between thermal or device offsets, and thus compensates for both. To avoid sampling the AC component of the ETF's output, the ETF is driven at $16f_{drive}$ during autozeroing. Since the ETF is a thermal LPF, this suppresses the AC component without affecting the DC component. When the s= 24μm ETF is driven at $16f_{drive}$ with f_{drive} = 42.67kHz, simulations show that the AC component of the first harmonic reduces by about 14.8x, which is close to 1st order roll-off behavior.

The main goal of the autozeroing circuit is to reduce ripple to a fraction of the signal swing. The absolute accuracy of the circuit is not very important, since any residual offset will still be modulated to f_{drive} by CH1 during normal operation.

The best time to autozero the g_m-stage is right before the start of a PDΣΔM conversion, so that it does not interfere with normal operation. The autozeroing caps needs to be sufficiently large so that

mismatch in the leakage of the minimum-size autozeroing switches does not significantly affect V_{AZ} during a conversion. $V_{AZ,CM}$ is mid-supply, so that the drain to body junctions are sufficiently reverse-biased and consequently the differential voltage does not greatly affect the leakage current. The sources of the autozeroing switches are connected to the output, which will be held to a virtual ground at mid-supply by the opamp while the gate voltage is zero, together minimizing the subthreshold leakage through the switches. Measurements show that V_{AZ} can be held to within ≈ 10mV for several seconds.

4.2.4 Integrator

The demodulated current is integrated using an active integrator based on a two-stage opamp. This ensures that the output of the folded cascode sees a low-impedance virtual ground with very little signal swing. The combination of a gain-boosted folded cascode and a two-stage opamp provides very high DC gain (>160dB over PVT).

The opamp schematic is shown in Fig. 4.6. The first stage consists of a folded cascode and the second stage is a common-source stage. Miller compensation guarantees stability (C_M = 1pF).

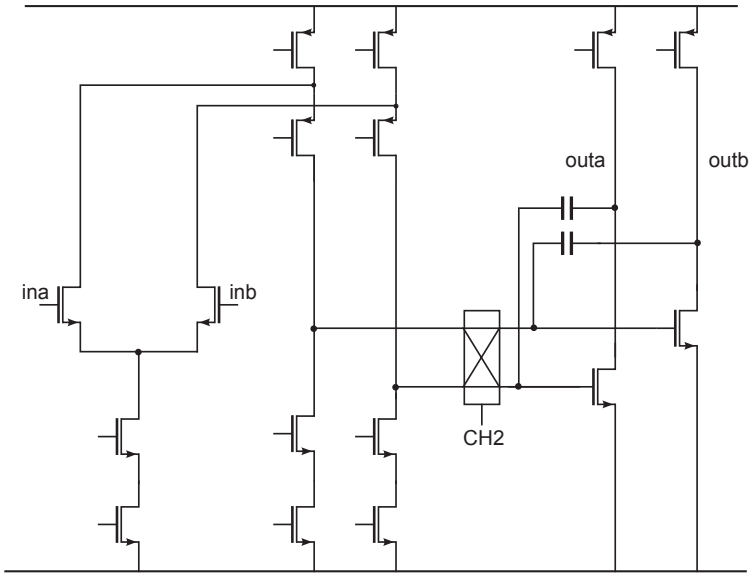


Fig. 4.6: Schematic of the two-stage opamp that implements the active integrator.

The bandwidth requirements of the opamp are determined by f_s. The PDΣΔM is a continuous-time ΣΔM, and thus a bandwidth of ≈$2f_s$ is sufficient. Since $f_{s,max}$ is only about 21kHz, the current consumption required is only several μA. This is only a fraction of the OTA bias currents and the current associated with P_{heat}.

Chopper CH2 reduces the effect of opamp offset; this will be discussed in section 4.3.

4.2.5 Comparator

The integrator output is sampled by a single-bit ADC, implemented by a latched comparator. The schematic is shown in Fig. 4.7 A differential pair converts the comparator's input signal to current, which is then mirrored to drive a positive feedback latch. When *eval* is low, the two SR-latch inputs are held at a diode from the positive rail and the latch is not set. At the rising edge of *eval*, positive feedback rapidly generates a digital output corresponding to the polarity of its differential input signal.

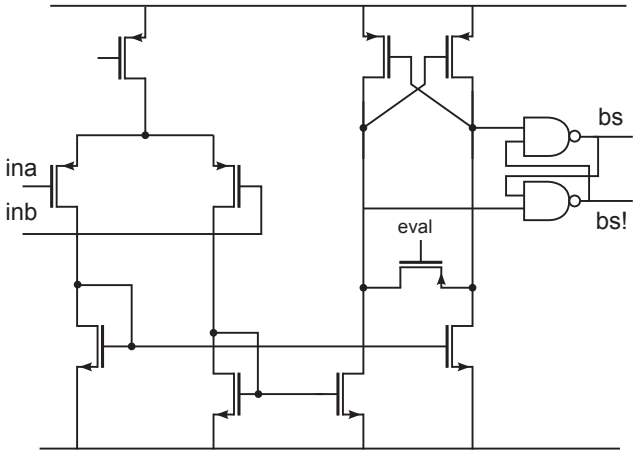


Fig. 4.7: Comparator schematic

The propagation delay is on the order of a few nanoseconds. Since f_s is low (and always $\leq f_{drive}$), this delay is small. The delay does not affect the feedback sampling time, since the phase DAC resynchronizes the bitstream.

4.3 PDΣΔM errors and error reduction methods

This section will discuss techniques that are used to ensure that ϕ_{ETF} is measured accurately. Sources of error can be classified into two groups: before the demodulator (chopper CH1 in Fig. 4.5), phase errors can add offset and spread to the measured ϕ_{ETF}; after the demodulator, DC currents have the same effect. These two types of error will be discussed separately.

4.3.1 Phase offset

An important source of error is associated with the finite delay between the ETF output and the demodulator. Both the thermopile's electrical filtering and the pole at the folding node of the g_m-stage low-pass filter the ETF's output signal on its way to the demodulator, which adds a nominal phase offset as well as potential phase spread to ϕ_{ETF}. This is illustrated by the schematic shown in Fig. 4.8. To meet the requirements outlined in Table 4.1, both filters should be sufficiently fast. A reasonably accurate quantitative estimate for their phase error, ϕ_{err}, is given by the phase shift of a simple 1st-order low-pass filter:

$$\phi_{err} = -\arctan\left(\frac{f_{drive}}{f_{-3dB}}\right) \tag{4.6}$$

For the absolute ϕ_{err} to be no more than 3mº at f_{drive}=42kHz, a -3dB-bandwidth of 800MHz is required. This translates into a minimum current in the output branches of the OTA: at e.g. a 100fF load capacitance at the folding node, this current should be on the order of 25µA. This is comparable to the current needed in the input pair to meet thermal noise requirements.

A more limiting constraint is associated with the filter at the OTA's input, which is composed of the thermopile's own RC network and the OTA's input capacitance, C_{in}. For R_{tp} = 16kΩ, C_{in} can be no more than 13fF, even when the thermopile's parasitic capacitances are ignored. This is unrealistically small.

The above requirements can be relaxed by tolerating a nominal ϕ_{err} and then designing the *spread* on ϕ_{err} to be at the target error level. To obtain an estimate for the spread on ϕ_{err} due to e.g. device mismatch and spread over process corners, ϕ_{err} was simulated over PVT. At f_{drive}=85kHz, ϕ_{err} changes from 144 to 185mº, so the total spread in ϕ_{err} is 41mº. This includes the ±20% variations on the thermopile resistivity and the parasitic capacitances, as well as (smaller) changes in the bandwidth of the OTA. At f_{drive}=42kHz, the phase error is ≈20mº over corners. Assuming a 10% variation within one corner, the spread in ϕ_{err} over one wafer lot will be below the target error level (3mº).

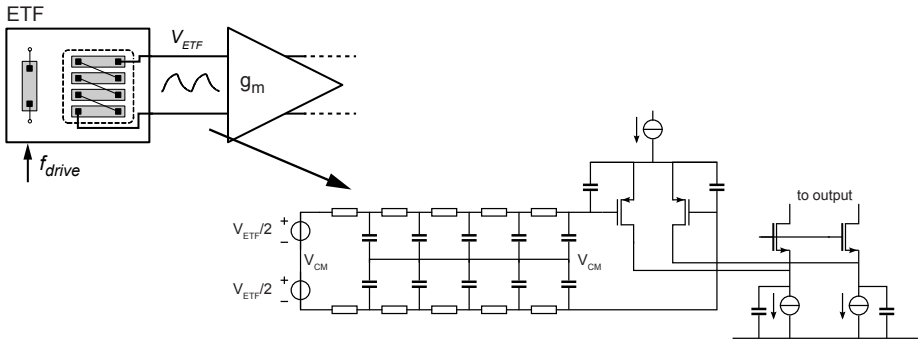


Fig. 4.8: Schematic showing the electrical filtering at the input of the g_m-stage.

When the design is pushed for high bandwidth, it should be noted that ϕ_{err} is only below the error threshold within a batch, and so it may be an appreciable source of wafer-to-wafer and batch-to-batch spread (41mº at f_{drive}=85kHz translates to ±0.1ºC).

A tradeoff exists between ϕ_{err} and resolution. Decreasing f_{drive} by a factor of N improves the ϕ_{ETF} / ϕ_{err} ratio by $\sqrt{N}$, but at very low f_{drive}, the ETF's intrinsic resolution degrades due to decreasing $d\phi_{ETF}/dT$. For the ETFs discussed in this work, a sufficiently wide range of f_{drive} delivers acceptable results. This is generally the case for larger ETFs.

However, for the scaled ETFs of future work, it will be more difficult to reduce the spread in ϕ_{err} to acceptable levels. For smaller s, the f_{drive} that yields the highest resolution will increase quadratically. The thermopile's RC filtering also reduces approximately quadratically, but the bandwidth requirement on the OTA will require it to consume much more current, making it much more challenging to achieve the same relative ϕ_{err}. Future work may employ other techniques such as ϕ_{err} calibration or cancellation through a form of phase-autozeroing.

4.3.2 Electrical offsets

4.3.2.1 G_m-stage residual offset

As discussed in section 4.2.3, the thermal V_{OS} and the g_m-stage offset are modulated to AC by the demodulating chopper, and the associated ripple is reduced by autozeroing. These error sources no longer contribute to errors in ϕ_{ETF}.

However, the g_m-stage will still have a residual offset associated with other circuit nonidealities around the demodulator. It is most useful to express this offset as an output-referred current, since DC errors at the output of the g_m-stage contribute directly to error in ϕ_{ETF}. Table 4.1 has shown that a 3mº phase error corresponds to an effective offset of no more than 7pA, and so the offset current at the g_m-stage output has a sensitivity of 0.48nA/ºC.

One source of residual offset is that of charge injection mismatch in the demodulator. If the charge required to establish the channel for one set of NMOS switches is different from the other pair, this difference gets injected into the folding node at every switching instant. The gain-boosted cascode

has a low input impedance, and thus the charge will flow into the cascode, and therefore into the output. The current associated with this mechanism is given by:

$$I_{OS,\Delta q_{inj}} = 2 f_{drive} \Delta Q_{inj} \tag{4.7}$$

The charge injection mismatch is a function of the total channel charge and the matching between the demodulating switches. Since Q_{inj} scales linearly with WL but the standard deviation of its relative spread (dQ_{inj}/ Q_{inj}) is inversely proportional to $\sqrt{}$(WL), minimum-size devices will have the lowest I_{OS}. In practice, switches that conduct a significant current may require too much voltage headroom when they are minimum size. In this design, switches with W/L = 3/0.18μm were used, having a total gate capacitance of about 14fF. For typical biasing voltages and assuming a 10% mismatch in charge injection, this leads to a 16fC charge being injected into the cascodes at every switching instant. For f_{drive} =85kHz, $I_{OS,\Delta qinj}$ is about 0.28nA. This translates into an error of about 0.6ºC, which is significantly larger than the 0.015ºC target: further offset cancellation methods are thus required.

A similar source of offset current is associated with the charging and discharging of parasitic capacitors at the source terminals of the chopper, i.e. where the differential signal current is injected into the output branch. Fig. 4.9 shows a basic folding node with a chopper at the sources of the cascode. The cascode transistors will also have offset, and so the differential voltage across the top terminals of the choppers will be $V_{OS,\text{casc}}$. Since the chopper switches at f_{drive}, the parasitic capacitors (C_p) at the folding nodes must see a square wave with amplitude $2V_{OS}$. The current spikes associated with the charging and discharging of C_p are rectified by the chopper and thus lead to an effective DC current at the chopper output. The amplitude of this current is given by:

$$I_{OS,Q_{CP}} = 4 V_{OS} C_p f_{drive} \tag{4.8}$$

For typical values of V_{OS} (10mV), C_p (50fF) and f_{drive} (85kHz), this leads to $I_{OS,\text{Qcp}}$ = 0.17nA, which is comparable to the chopper's charge injection and thus requires mitigation.

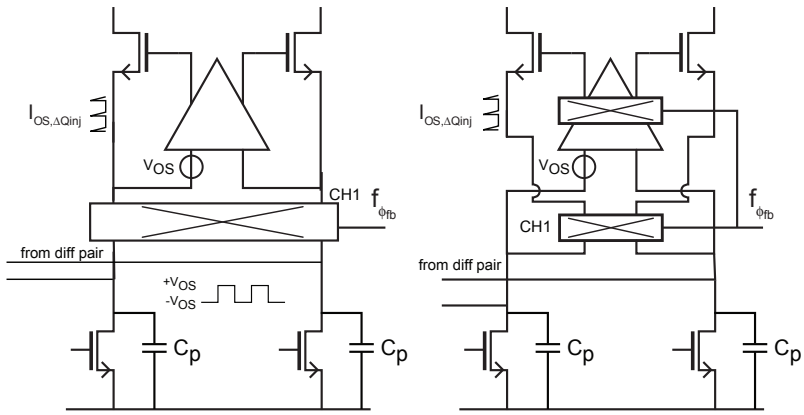


Fig. 4.9: Folding node choppers with nonidealities; the circuit on the right prevents the charging and discharging of C_p.

An elegant way to prevent the charging and discharging of these parasitic caps is by forcing a virtual ground at the folding node using the n-side gain-boosting amplifier [4.3]. This is also shown in Fig. 4.9. The demodulator is still located at the cascode sources, but the amplifier inputs are connected at the drains of the current sources instead of after the demodulator. The amplifier now holds these (high-capacitance) nodes at a constant virtual ground, so no displacement current flows into C_p. To

maintain the correct loop polarity, the output of the amplifier is also chopped. This circuit effectively prevents $I_{OS,Qcp}$, but does not mitigate $I_{OS,\Delta qinj}$.

A possible alternative to lower the offset is reducing f_{drive}. Operating at low f_{drive} increases signal amplitude while linearly reducing I_{OS}. However, f_{drive} has a lower bound due to SNR requirements, and thus a more effective method to deal with these currents is required, which will be described in section 4.3.2.4.

4.3.2.2 Opamp offset

The input-referred offset of the opamp can also lead to a DC error current. Fig. 4.10 shows this schematically:

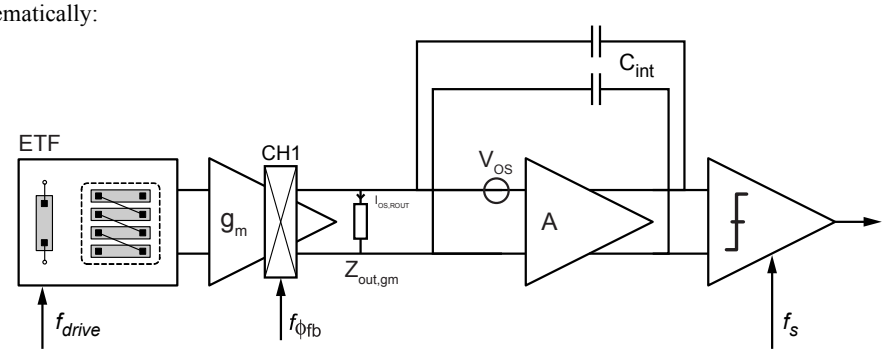


Fig. 4.10: Opamp offset leading to an error current due to the finite output impedance of the g_m -stage.

During normal operation, the differential voltage across the opamp inputs will be equal to $-V_{OS}$, plus a small voltage associated with the finite DC gain of the opamp. From Table 4.1, the input-referred V_{OS} of the readout circuit should be no more than 12nV. When the opamp's V_{OS} is 10mV, this requires a DC gain of at least 118dB. Although gain boosting greatly helps to increase R_{out}, the lower intrinsic gain of MOS devices in scaled technologies may reduce the DC gain below the required level, in which case this offset should be suppressed.

4.3.2.3 Comparator offset

The comparator offset is on the order of 10mV. Since this offset is located at the integrator output, it will be noise-shaped by the ΣΔM. After a single conversion of N samples, the effective demodulator-referred DC error current $I_{OS,comp}$ is given by:

$$I_{os,comp} = \frac{V_{OS,comp} C_{int}}{N t_s} \tag{4.9}$$

For $V_{OS,comp}$ = 10mV, C_{int} = 40pF, N = 32768 and $f_s = f_{drive}/4$ = 32kHz, this error is about 0.4pA, which is negligible.

4.3.2.4 System-level chopping

One way to reduce the effect of residual offset currents is to add a layer of low-frequency choppers around the DC error sources in the circuit. Fig. 4.11 shows where these choppers can be placed:

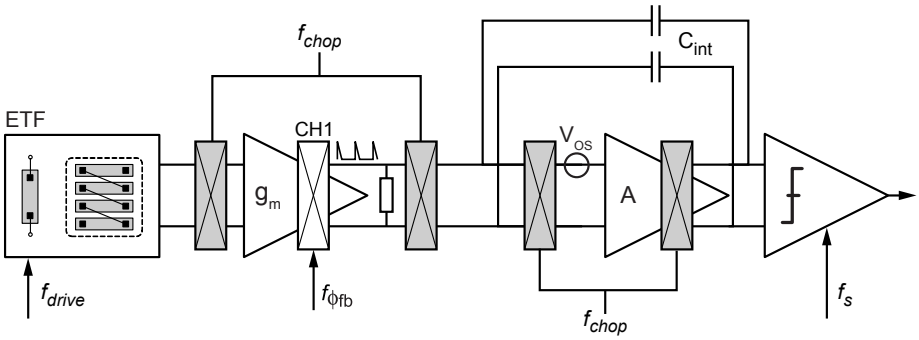


Fig. 4.11: Chopping the charge injection of CH1 and the opamp offset.

By periodically reversing the polarity of these switches by toggling f_{chop}, both the charge injection mismatch and the opamp's V_{OS} will be converted to AC signals, so that they will be filtered by C_{int} and by the decimation filter of the ΣΔM. Although these choppers will also have charge injection mismatch, they can be operated at much lower frequency, so that the corresponding error current is small. Typical values are $f_{chop}=f_s/8192$ or $f_{chop}=f_s/16384$, corresponding to only one or two f_{chop} periods per conversion. The associated low-frequency tone will be at a notch of the decimation filter, so it will not deteriorate the PDΣΔM's resolution.

The chopping as shown in Fig. 4.11 can be simplified in two ways. Firstly, the chopper at the input of the g_m-stage can be shifted through the ETF: phase-shifting P_{heat} by 180º implements a polarity reversal [4.12]. In this way, no switches are needed at the sensitive ETF output. Secondly, the two choppers at the opamp virtual ground can be combined into a single chopper in the opamp's feedback path. This is shown below in Fig. 4.12:

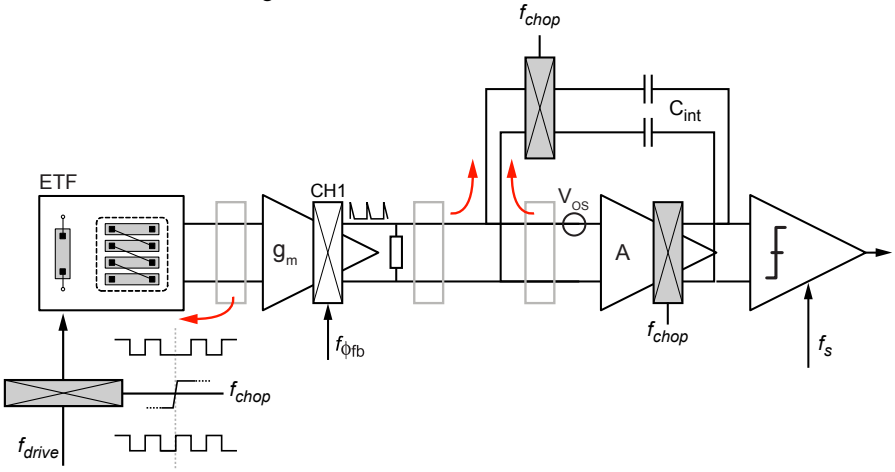


Fig. 4.12: Simplifying the chopping implementation.

Over the military temperature range and for typical device mismatch, the residual offset was simulated to be less than 7.5pA, which is close to the 7pA target. Therefore, the architecture is sufficiently precise to enable ETF characterization.

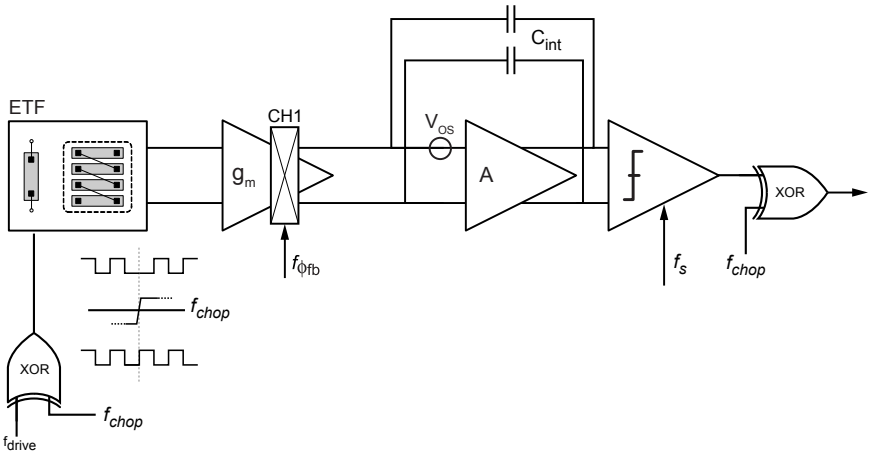


Fig. 4.13: Alternative system-level chopping topology.

A second way of implementing system-level chopping is by combining two incremental PDΣΔM conversions, each corresponding to a different state of f_{chop}. In this way, the system can be simplified further, as shown in Fig. 4.13. Polarity-reversing switches are needed at the heater drive and at the comparator output, and they are implemented by simple XOR gates. This architecture has the advantage of also cancelling the offset of the comparator, but has the disadvantage of increased quantization, because two decimated PDΣΔM outputs are digitally combined. The signal chain should be sufficiently linear to ensure that the combined digital outputs of two conversions are equal to the offset-free output, i.e:

$$\frac{D_{OUT}(I_{SIG}+I_{OS})+D_{OUT}(I_{SIG}-I_{OS})}{2}=D_{OUT}(I_{SIG}) \tag{4.10}$$

This is easier to achieve when the residual offset before system-level chopping is low. The chopped gain booster topology, in combination with careful layout of the demodulating switches, typically achieves a sufficiently low residual offset to enable this method of system-level chopping.

Although the signal path is highly linear for DC offset currents, the PDΣΔM exhibits a fundamental nonlinearity in its ϕ_{ETF} to I_{DC} transfer function. This will be discussed more extensively in the next section.

4.4 Nonlinearity of a 1-st order PDΣΔM

The PDΣΔM is only fully transparent if it is also fully linear, i.e. there should be a linear relationship between ϕ_{ETF} and the digital output. Unfortunately, this is not the case for the first-order single-bit PDΣΔM discussed in the previous section. Consider the phase detection node, in which the phase-shifted output signal V_{ETF} is multiplied by a feedback signal (ϕ_{fb}) having one of two reference phase shifts (ϕ_0 or ϕ_1). The DC output of the phase detector is given by:

$$dc=\frac{4}{\pi^2}\cos(\phi_{ETF}-\phi_{fb}) \tag{4.11}$$

Because of the cosine term, this relationship can only be considered linear near $dc \approx 0$, which holds around $(\phi_{ETF}-\phi_{fb}) = 90°$. Since this condition is only met by the *average*, but not by the *instantaneous* values of ϕ_{fb}, the transfer function of the PDΣΔM is nonlinear. This is also illustrated in Fig. 4.14:

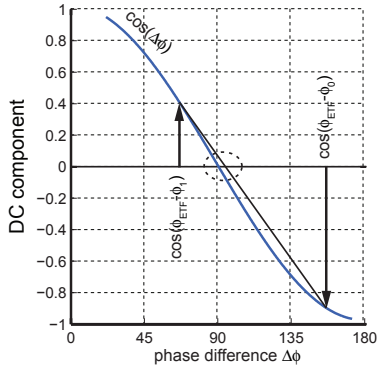


Fig. 4.14: Phase detector cosine nonlinearity; the straight line between the two ΣΔM error signals does not cross zero at cos($\Delta\phi$)=0.

After a large number of samples, the average DC component at the input of the integrator will be approximately zero. This means that the bitstream average μ must satisfy:

$$\mu\cos(\phi_{ETF}-\phi_1) = -(1-\mu)\cos(\phi_{ETF}-\phi_0) \tag{4.12}$$

The bitstream average μ is then given by:

$$\mu = \frac{\cos(\phi_{ETF}-\phi_0)}{\cos(\phi_{ETF}-\phi_0)-\cos(\phi_{ETF}-\phi_1)} \tag{4.13}$$

After a conversion, μ can be used to calculate ϕ_{out}, the digital approximation of ϕ_{ETF}. The fraction μ defines a point on the phase span defined by ϕ_0 and ϕ_1. Because (4.11) is only zero for ($\phi_{ETF} - \phi_{fb}$) = 90º, this point has to be shifted by 90º to obtain the following expression for ϕ_{out}:

$$\phi_{out} = \mu\phi_1 + (1-\mu)\phi_0 + 90^\circ \tag{4.14}$$

In Fig. 4.15, ϕ_{out} is plotted as a function of ϕ_{ETF} for a PDΣΔM with (ϕ_0, ϕ_1) = (-45°, 45°), together with its relative nonlinearity. It can be seen that a single-bit PDΣΔM has a systematic nonlinearity that is determined by the choice of its phase shift references.

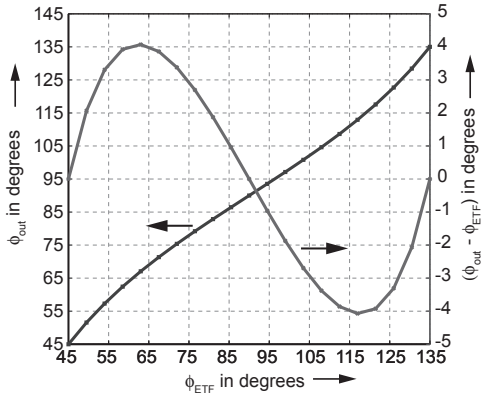


Fig. 4.15: PDΣΔM transfer function and corresponding linear error.

Because the PDΣΔM nonlinearity is systematic (and thus insensitive to e.g. PVT variations), it does not necessarily introduce temperature-sensing errors: for a given μ and known phase references ϕ_0 and ϕ_1, the ϕ_{ETF} that was digitized can be determined by inverting or "unwrapping" Eq. (4.13) to obtain the following equation:

$$\phi_{ETF} = \tan^{-1}\left(\frac{\mu\cos(\phi_1-\phi_0)-\mu+1}{\mu\sin(\phi_1-\phi_0)}\right)-\phi_0 \tag{4.15}$$

For the case in which (ϕ_0, ϕ_1) is (-45º, 45º), this reduces to:

$$\phi_{ETF} = \tan^{-1}\left(\frac{\mu+1}{\mu}\right)+\frac{3\pi}{4} \tag{4.16}$$

Although the nonlinearity of a single-bit PDΣΔM is systematic and thus can be digitally corrected for, these rather complex trigonometric equations requires significant computational overhead.

For some applications, such as the ratiometric temperature sensor discussed in section 3.5, a more linear phase-to-digital converter is required. Fig. 4.14 shows that the PDΣΔM is nonlinear because of cosine distortion. However, for values close to 90º, the cosine function is near-linear. The figure also shows that reducing the distances $|\phi_{ETF}-\phi_0|$ and $|\phi_{ETF}-\phi_1|$ would reduce the deviation to $\cos(\phi_{ETF})$, thus improving linearity. It can be shown that the worst-case nonlinear error as a function of $|\phi_1-\phi_0|$ is given by:

$$\left|\phi_{out}-\phi_{in}\right|_{\max} \approx \frac{\left|\phi_1-\phi_0\right|^3}{e^{12.2}} \tag{4.17}$$

Linearity shows a cubic relationship to the phase input range, so a PDΣΔM is very linear for small $|\phi_0-\phi_1|$. Unfortunately, such small phase input ranges also imply that the useful temperature range is very limited. In a single-bit PDΣΔM, the input range and its nonlinearity are strongly coupled. To obtain both linearity and wide phase range, a different architecture is required.

4.5 Zoom-in phase-to-digital converter (ZPDC)

One solution to the cosine nonlinearity problem is the use of a multi-bit PDΣΔM; several topologies were presented in section 3.3.3.3. The cosine nonlinearity of a multi-bit PDΣΔM is determined by the spacing between feedback levels, rather than the full scale input range. Typical multi-bit architecture require both an N-bit phase DAC as well as an N-bit ADC, increasing system complexity.

Section 3.3.3.3 also presented the *zoom-in* phase-to-digital converter, or ZPDC. In a ZPDC, ϕ_{ETF} is first coarsely quantized by using a fast, low-resolution phase digitizer. The second step then uses a PDΣΔM to obtain a high resolution. Since the ZPDC will be discussed in more detail in this section, the block-level diagram is repeated for clarity:

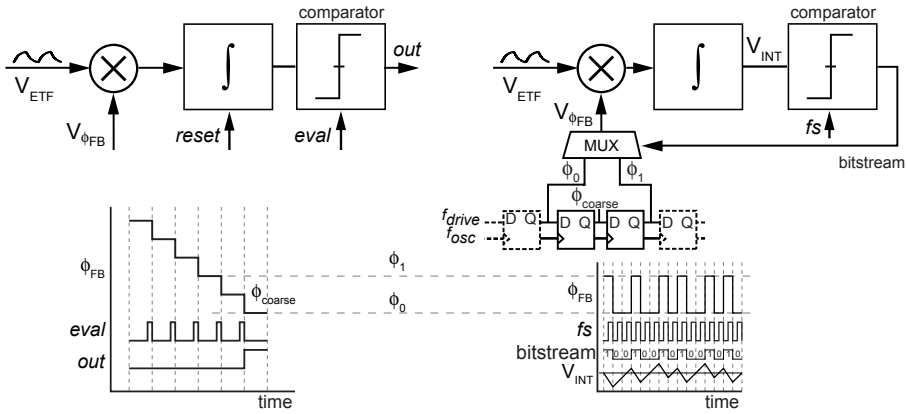


Fig. 4.16: Zoom phase-to-digital converter (ZPDC). The left hand side shows the first coarse phase digitization step, the right hand side shows the second phase ΣΔM conversion.

The ZPDC's coarse phase executes a single-slope digitization algorithm. During this process, the ETF's output signal is multiplied with another signal at f_{drive} whose phase shift, ϕ_{ramp}, starts at 45° (the original ϕ_1) and is ramped downwards in ϕ_{step} decrements. For each step, the multiplier output is integrated for 32 sampling periods, after which a sampled comparator detects the polarity of V_{int}. Since the DC component of the multiplier is approximately proportional to $cos(\phi_{ETF} - \phi_{ramp})$, the comparator detects a zero crossing when $\phi_{ETF} - \phi_{ramp} \geq 90°$. At this point, the instantaneous value of ϕ_{ramp} is stored as ϕ_{coarse}, and the integrator is reset.

In the second digitization step, two phase references ϕ_0, $\phi_1 = \phi_{coarse} \pm \phi_{step}$ define the input range of a 1st-order, single-bit PDΣΔM, whose decimated output ϕ_{fine} represents a weighted average of ϕ_0 and ϕ_1. Since ϕ_{ETF} is already coarsely known, the PDΣΔM's input range can be significantly reduced, from ±45º to e.g. only ±2.8125°. This reduces its quantization error by 16x, and reduces the cosine nonlinearity associated with the multiplying phase detector by $\approx 16^3$ = (4096) times, thus enabling accurate phase detection. The fine phase outputs ϕ_{out}, which now has an inaccuracy of better than 16 bits: errors due to nonlinearity are at the noise level. Since only the result of the fine phase is used, errors in ϕ_{coarse} do not appear in the end result.

The minimum step size, ϕ_{step}, by which the total phase range can be divided, is determined by the ratio f_{ref}/f_{drive}, in which f_{ref} is a high reference frequency that is used to synchronously delay f_{drive}; in the example above, f_{ref} =6.4MHz and f_{drive} = 50kHz, so that the minimum ϕ_{step} is 2.8215°. Although for scaled ETFs in microprocessors, f_{drive} will increase to several MHz, frequencies of several GHz are available to generate sufficiently small phase steps.

Care has to be taken to ensure that ϕ_{ETF}-ϕ_{coarse} is always within the PDΣΔM input range. The first phase outputs ϕ_{coarse} if $\phi_{coarse} < \phi_{ETF} < \phi_{coarse} + \phi_{step}$, and so ϕ_{ETF} could be very close to $\phi_{coarse} + \phi_{step}$, which is very close to the upper input level of the PDΣΔM that follows. In the presence of offset, the first step might mistakenly choose a ϕ_{coarse} for which the PDΣΔM will clip. To remedy this situation, the first step is operated by synchronizing ϕ_{fb} to the negative clock edge, effectively creating thresholds at $\phi_{coarse} \pm \phi_{step}/2$. This ensures that ϕ_{coarse} will always be $\phi_{step}/2$ away from the limits of the PDΣΔM input range.

The effect of offset and 1/f noise is reduced in the same way as for the PDΣΔM. Typically, only the second (ΣΔM) phase requires offset reduction techniques. The accuracy of the first step only has to be at the ϕ_{step} level.

The total conversion time of the fine conversion is still limited by thermal noise, and thus remains mostly unchanged from the single-bit PDΣΔM. For equal P_{heat}, thermal noise and conversion rate, the ZPDC actually has a slightly higher resolution. This is because in a single-bit PDΣΔM, the amplitude of the phase error current after cosine distortion is slightly lower than for the linear phase detector, depending on $|\phi_{ETF}-\phi_{REF}|$ [4.6].

Care also has to be taken that f_{drive} does not change between fine and coarse conversion. The ZPDC operates under the assumption that its input remains relatively constant, but ϕ_{ETF} is a function of both T and f_{drive}. For the self-referenced temperature sensor, this means that the on-chip oscillator should have reasonable supply and temperature stability. ETF-based thermal management sensors in a μP should be driven by a clock that is not subjected to e.g. throttling, or the ZPDC conversion should be reset after a step in f_{drive}.

4.5.1 Circuit implementation

The ZPDC can effectively re-use the building blocks of the PDΣΔM to implement both the coarse and the fine phase. In the coarse phase, the integrator is still required to average the demodulated output current, for which it operates as a time-dependent gain block. A switch across C_{int} is added to reset the integrator at the start of each coarse digitization step. The comparator is used to decide on the polarity of the integrator output, as in the PDΣΔM.

During the fine conversion, the PDΣΔM operates as for the non-zooming architecture, but with greatly reduced phase input range. Since the phase references are closer to ϕ_{ETF} (by zooming factor N), the demodulated error currents are proportionally smaller. Advantageously, this means that the value of C_{int} can also be reduced by a factor N, saving area and (potentially) power consumption.

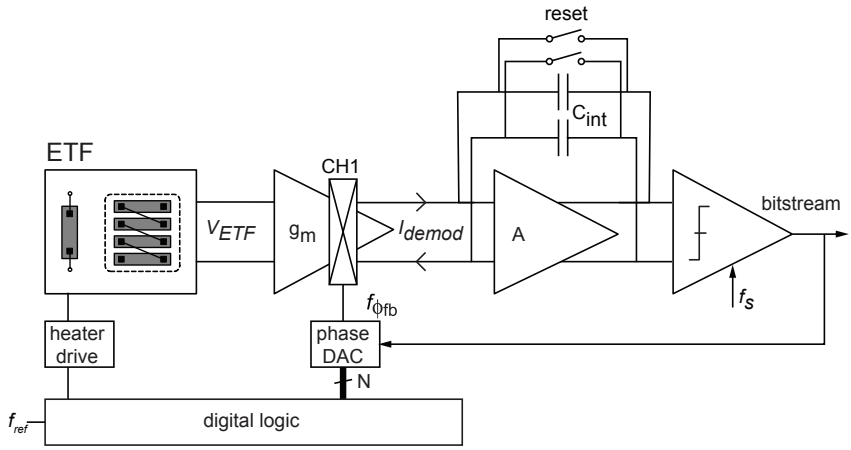


Fig. 4.17: ZPDC implementation; during the coarse phase, the integrator is reset after each conversion.

Fig. 4.17 shows the PDΣΔM circuitry as configured for two-step operation. Some extra digital logic is required to generate the two-step timing and the outputs of the coarse and fine conversion are digitally combined to provide D_{out}(T). The most straightforward way to implement the phase DAC is to use a synchronous divide-by-N counter, which will output a ϕ_{fb} that can be stepped by ϕ_{step} by incrementing the starting value of the counter by one. The simplest way to run the first step of the ZPDC is then to perform a single slope algorithm by increasing ϕ_{fb} until the comparator triggers. Successive approximation is much faster (N vs. $\log_2$(N)) and has a temperature-independent conversion time, but additional digital logic is required. Since the number of samples needed in the second step is much larger due to thermal noise requirements, the conversion time of the first step is typically not critical.

Fig. 4.17 also shows that the analog circuit overhead is very small; the large reductions in nonlinearity and quantization noise greatly outweigh the overhead, making the ZPDC attractive for ETF readout. As will be shown in the next section, the ZPDC can also help to linearize the ϕ_{ETF}(T) transfer function.

4.6 Linearization

The previous section has shown that the cosine nonlinearity of the ZPDC is orders of magnitude smaller than for a 1-bit PDΣΔM. However, a linear ETF-based temperature sensor not only requires a linear PDC; the $T^{0.9}$ temperature dependence of D_{Si} also needs to be corrected for: over the -55ºC to 125ºC range, this leads to ±5ºC nonlinearity. This section will discuss several methods to linearize single-ETF temperature sensors.

4.6.1 Method 1: PDΣΔM phase reference adjustment

The first linearization method is based on adjusting the cosine nonlinearity of the 1-bit PDΣΔM in such a way as to compensate the $T^{0.9}$ nonlinearity from D_{Si}. The PDΣΔM's cosine nonlinearity can be tuned by placing phase references ϕ_0 and ϕ_1 in such a way so as to map ϕ_{ETF} onto a part of the PDΣΔM's transfer function for which the total transfer function (from temperature to digital output) is most linear. In other words, linearizing the following equation:

$$\mu(T) = \frac{\cos(s\sqrt{\frac{\pi f_{drive}}{D_{Si}(T)}} - \phi_0)}{\cos(s\sqrt{\frac{\pi f_{drive}}{D_{Si}(T)}} - \phi_0) - \cos(s\sqrt{\frac{\pi f_{drive}}{D_{Si}(T)}} - \phi_1)} \tag{4.18}$$

The most straightforward way to optimize for linearity using this methods is to perform a 2D-sweep of ϕ_0 and ϕ_1 and perform a linear fit on $\mu(T)$. Fig. 4.18 shows the resulting maximum linear error over the -40ºC to 105ºC range. A relatively weak optimum exists, and so several ϕ_0, ϕ_1 combinations give acceptable results.

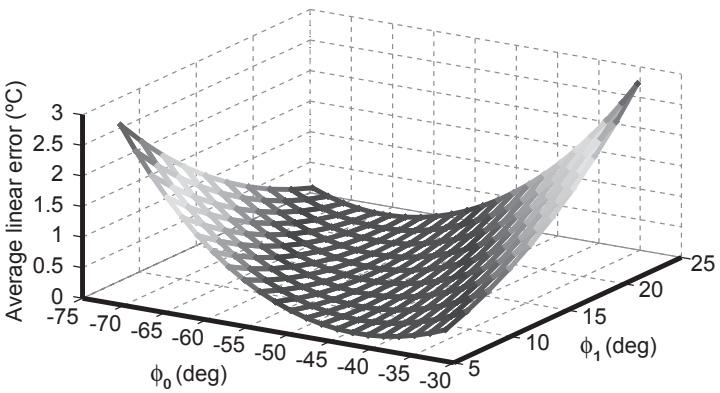


Fig. 4.18: Nonlinearity as a function of ϕ_0 and ϕ_1.

Fig. 4.19 shows the measured ϕ_{out} (=μ(T) scaled by $|\phi_1 - \phi_0|$) and the linear error as a function of temperature. The results show that ±0.7°C nonlinearity can be achieved over a 150°C range [4.5].

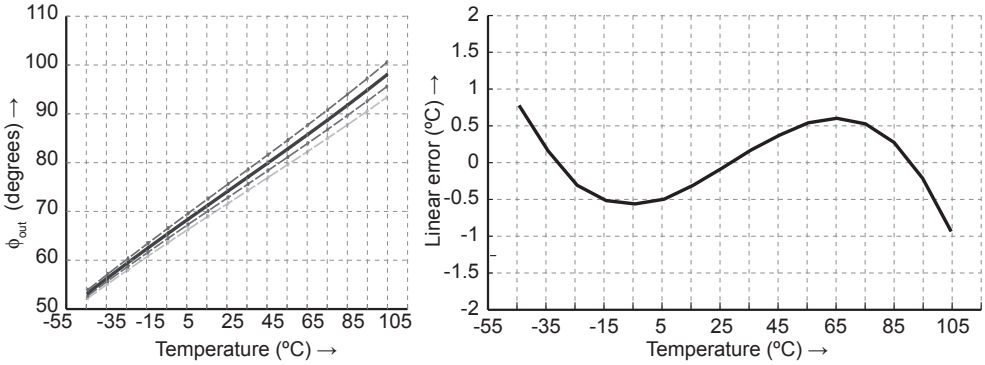


Fig. 4.19: Measured ϕ_{out} and linear error over temperature.
The dashed transfer functions correspond to neighboring ϕ_0, ϕ_1 settings.

The main advantage of this method is that no extra analog or digital overhead is required for an improvement in linearity; a simple 1-bit PDΣΔM is sufficient. In the above example, linearity was optimized over a relatively wide temperature range. For thermal management applications, linearity could be optimized for a more narrow range of around 30ºC to 100ºC. In this way, 0.1ºC – 0.2ºC nonlinearity is achieved for temperatures critical to self-heating, while a larger error is tolerable at low temperature.

There are some disadvantages to this method. Because the two nonlinear effects are mathematically quite different, the linear error rapidly increases when the range is extended beyond the Fig. 4.19 example. Additionally, optimum cancellation takes place when ϕ_{ETF} covers about one third of the total PDΣΔM input range; this implies that the quantization noise is 3x higher than necessary. A final requirement is the need for a high frequency f_{ref}, since ϕ_0 and ϕ_1 can only be stepped in f_{ref}/f_{drive} steps.

4.6.2 Method 2: Zoom PDC-based linearization

When the zoom PDC is used, the result of the coarse phase can be used to linearize the system. Fig. 4.20 shows a typical ETF's measured phase and temperature errors to a linear fit. The vertical lines in the phase plot (on the left) indicate the size of each of the coarse phase bins: the ZPDC will have a fixed ϕ_{coarse} inside each of these bins. As can be seen, the static $T^{0.9}$ nonlinearity associated with D_{Si} is much more bounded *within* each bin than over the full range. This feature can be used to linearize the temperature sensor. Using a lookup table, a unique digital phase adjustment can be selected for each ϕ_{coarse}, i.e. $\phi_{adj}(\phi_{coarse})$. During the fine conversion, the heater signal is then phase-shifted by ϕ_{adj}, so that the final conversion result is more linear. As for method 1, this works because the ETF's nonlinearity is systematic and well-defined.

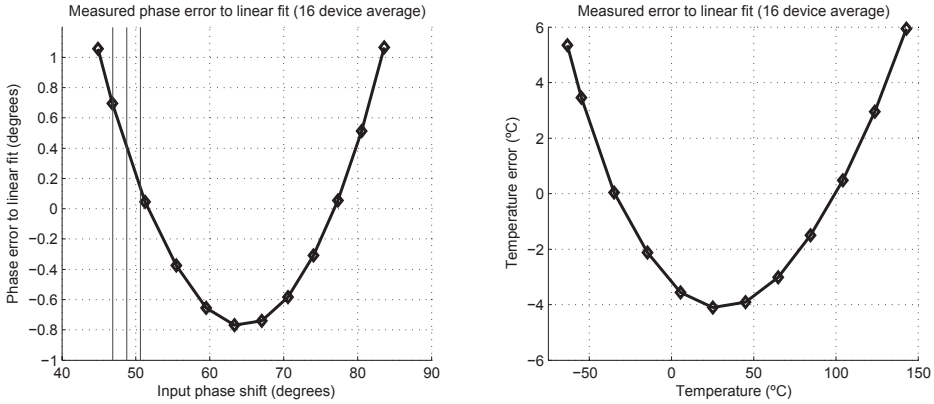


Fig. 4.20: Measured phase and temperature nonlinearity for a typical ETF.

One way to apply ϕ_{adj} to the heater driving signal is by duty-cycle modulation, as shown in Fig. 4.21. After having decided the coarse phase bin in step one, a LUT selects an inverse linear phase error to be added to ϕ_{ETF} by modulating the duty cycle of P_{heat}. Since many heat pulses are part of a conversion and the on-time of each pulse can be controlled in f_{ref}/f_{drive} steps, the trimming resolution is on the order of $f_{ref}T_{conv}$, which is very high (≈10e6).

Reducing the heater on-time from a duty cycle δ=50% to e.g. δ=49% can be seen as adding a narrow pulse with on-time δ=1% and negative amplitude $2P_{heat}$ to the heater signal. After demodulation and integration, this effects an approximately -7.2º phase shift.

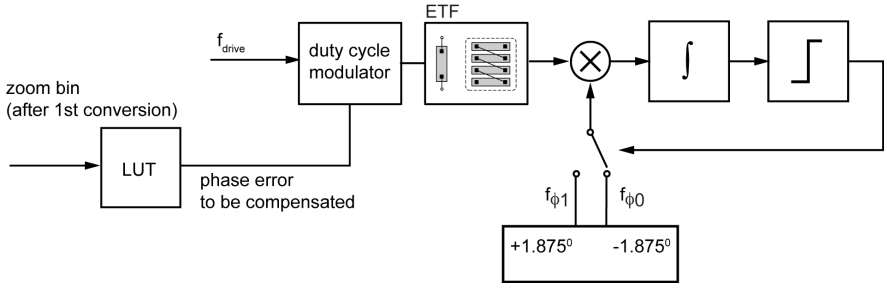


Fig. 4.21: LUT-driven duty cycle modulator adjusts nonlinearity of ϕ_{ETF} based on the first zoom step.

Since only one phase compensation value can be defined for each ϕ_{coarse} bin, the residual $T^{0.9}$ nonlinearity within a bin will limit the linear error.

This method was tested on a typical silicon ETF read out by the ZPDC. Fig. 4.22 shows the average temperature error for 16 devices over the -70ºC to 150ºC range; The LUT required 20 14-bit entries. The resulting nonlinearity is no more than ±0.2ºC, showing the efficacy of this method.

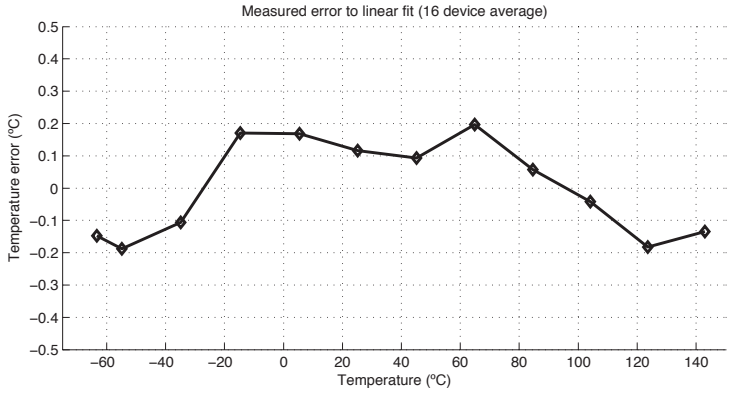


Fig. 4.22: Measured nonlinearity after applying duty-cycle correction.

4.6.3 Method 3: digital linearization

A third way of linearization is to filter ϕ_{ETF} in the digital domain instead of in the PDC. This can be done by either a non-linear decimation filter, or by processing the decimated output with a nonlinear digital scaler. The former has been implemented for BJT-based temperature sensors in the form of a piece-wise linear decimation filter [4.2], reducing a nonlinearity of up to 2ºC to near-zero. The latter is a more brute force method that would require significant digital overhead in the form of a large lookup table. This method was not considered in this work.

From simulations, a 4-segment piecewise linear decimation filter reduces nonlinearity significantly. Fig. 4.23 shows the simulated result; the resulting nonlinearity is well below ±0.1ºC, which is sufficient for many applications.

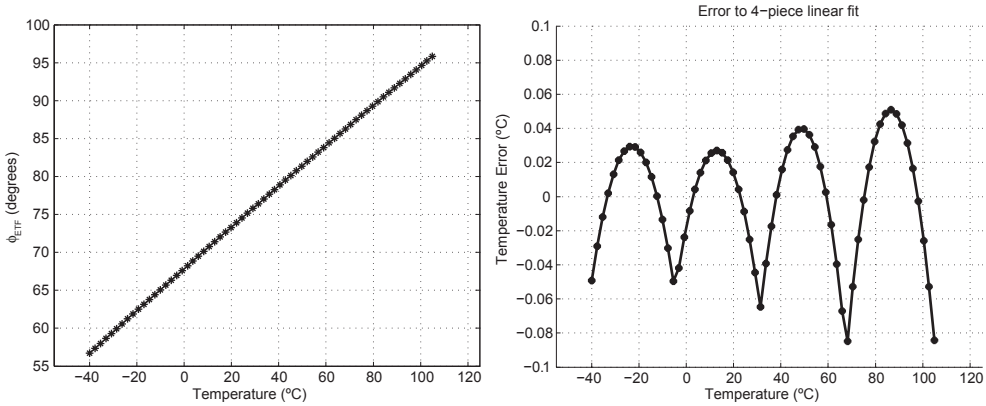


Fig. 4.23: Simulated linearization using a 4-segment piecewise linear decimation filter.

Linearization in the digital backend was not explored in further detail, in part because methods 1 and 2 achieve good linearity. However, it could be of future interest for thermal management applications, since digital computing power is abundant on modern microprocessors.

4.6.4 Multi-ETF temperature sensors

As was already presented in section 3.5.3, the nonlinearity of a self-referenced sensor can be minimized by scaling s_1 and s_2, i.e. the heater-sensor spacing of each of the two ETFs. This yields a near-linear output, having a nonlinear error of about ±0.5ºC over the -55ºC to 125ºC range.

Methods 2 and 3 can be used to further improve linearity as required. When a phase offset is added to f_{drive}, both ETFs should be driven at the same offset, since otherwise ϕ_{ox} will be incorrect.

4.7 PDΣΔM measurements & characterization

In order to prove that PDΣΔMs are an effective tool to digitize ϕ_{ETF}, their performance is first established separately from the ETF. To this end, a PDΣΔM with a 1/1000 resistive divider input instead of an ETF was implemented in a 0.18µm CMOS test chip. The chip has a pad-limited area of $2mm^2$, with the PDΣΔM occupying $0.18mm^2$. The PDΣΔM draws about 280µA from a 1.8V supply. A chip micrograph can be seen in section 5.1, although in this case the ETF was replaced by the aforementioned resistive divider. The total resistance of the divider was only about 5kΩ, to ensure that the RC-filter created by the divider and the input capacitance of the g_m-stage adds only negligible phase shift at f_{drive}.

Fig. 4.24 shows a schematic of the test setup. A discrete function generator was used to drive an off-chip single-ended to differential converter so as to drive the divider with $1V_{pp}$ to $100mV_{pp}$ differential signals, corresponding to typical P_{heat} levels at V_{TEST}. The input common-mode was set to the drop of a diode-connected NMOS device. The PDΣΔM operates as for an ETF, including the same offset reduction and filtering techniques as discussed above. An accurate phase relationship between V_{TEST} and $f_{\phi fb}$ was ensured by triggering the function generator on the same rising edge that drives the on-chip synchronization circuitry.

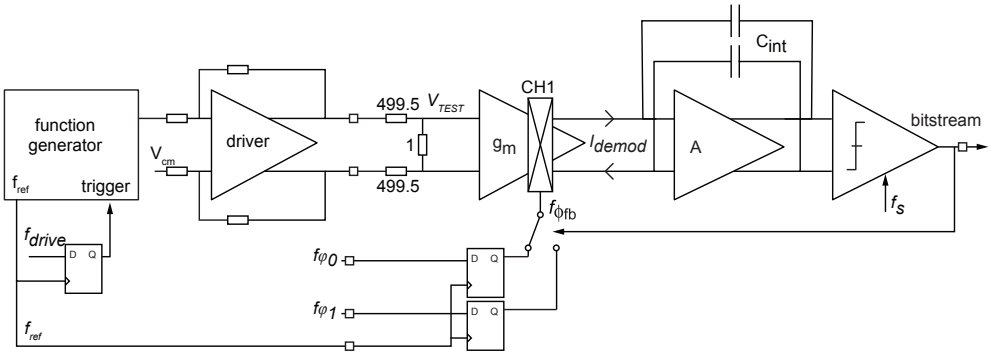


Fig. 4.24: Test setup for PDSDM characterization. The bondpad squares indicate the chip boundary.

The PDΣΔM's output was measured for both sine wave and square wave input signals, for several devices and over temperature. The input phase shift was swept in 1º steps; the PDΣΔM had a phase input range of 90 degrees and was operated at f_{drive} = 85.33kHz and f_s = 2.67kHz, and 32768 samples were taken per data point.

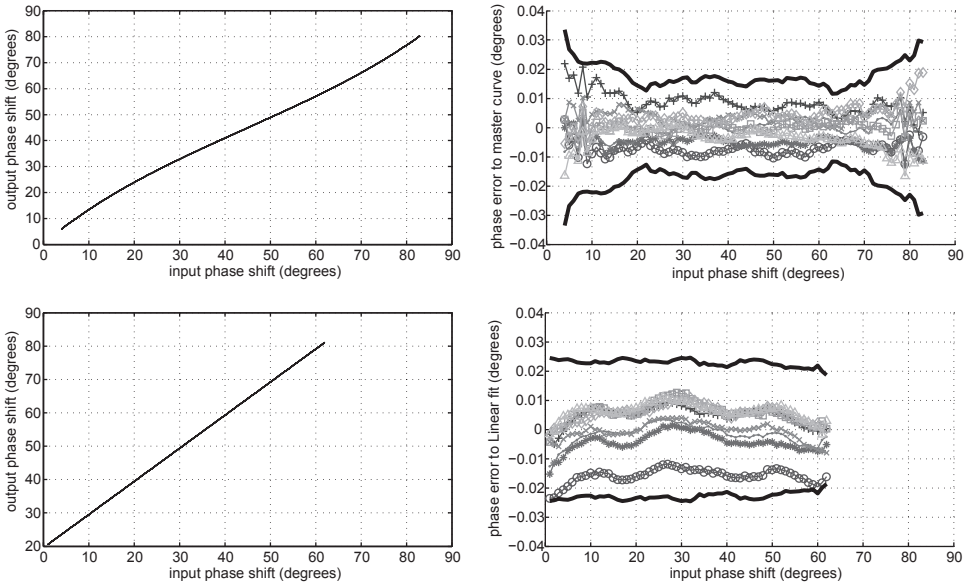


Fig. 4.25: Measured PDΣΔM transfer function and DC error for sine wave and square wave inputs.

Fig. 4.25 shows the PDΣΔM's measured digital output for 8 devices driven by a sinewave (top) and a square wave (bottom) at f_{drive} = 85.33kHz. It also shows the error of individual devices over the phase input range. For the sinewave input, the PDΣΔM exhibits the predicted nonlinearity, while it is linear for a square wave.

For a sinewave input, the phase error of the PDΣΔM is about ±0.018mº (3σ). For an s=24µm ETF driven at f_{drive} = 85.33kHz, this corresponds to a temperature error of about ±0.07ºC (3σ). Each error source was designed to contribute a phase error of only 4mº, so this is somewhat higher than expected; this could be due to modeling inaccuracy, timing error or crosstalk. Fortunately, the PDΣΔM is still sufficiently precise for ETF characterization. Moreover, there are no sources of nonlinearity other than the systematic cosine distortion, showing that the ETF's transfer function will be accurately measured.

When driven by a square wave, the PDΣΔM actually operates as a linear delay-to-digital converter; from the bottom right graph in Fig. 4.25, linearity is at the 5mº level, corresponding to about 14 bits. The linearity of the system without the PDΣΔM was not established, so it is not clear if this nonlinearity is due to the PDΣΔM or due to other error sources (such as the waveform generator).

For square-wave inputs, the PDΣΔM also has a slightly higher phase error. This can be attributed to the higher-order harmonics of the input signal, since their phase shift will spread more strongly with spread in the bandwidth of the g_m-stage. In normal operation, these harmonics are pre-filtered by the ETF.

Due to limitations in the test setup, the phase input range does not extend to the full 90 degrees available to the PDΣΔM. The waveform generator uses direct digital synthesis to 'draw' the PDΣΔM input signal; the starting point of this waveform was accurately controlled by a trigger signal (from the PDΣΔM timing logic), but the endpoint was not accurately known; the waveform generator needed a certain internal setup time to prepare its output for the next trigger. This means that for the above data, the effective f_{drive} was slightly *higher* than indicated; the results above have poor absolute accuracy, but they do show that the variations between several devices are small. This shows that PDΣΔM error will not increase the device-to-device temperature error of the devices under test.

4.8 Concluding remarks

This chapter has presented the design of a precision PDΣΔM for ETF characterization. From the results above, it can be concluded that the PDΣΔM is sufficiently accurate for ETF readout. This means that the measurement results, as presented in the next Chapter, can be reliably attributed to ETFs.

4.9 References

[4.1] K.A.A. Makinwa and M.F. Snoeij, "A CMOS temperature-to-frequency converter with an inaccuracy of ±0.5°C (3s) from –40 to 105°C," *IEEE J. Solid-State Circuits*, vol. 41, is. 12, pp. 2992 – 2997, Dec. 2006

[4.2] M.A.P. Pertijs and J.H. Huijsing, Precision Temperature Sensors in CMOS Technology, Springer, 2006.

[4.3] S.M. Kashmiri et al., "A Temperature-to-Digital Converter Based on an Optimized Electrothermal Filter", *IEEE J. Solid-State Circuits*, vol. 44, no.7, pp. 2026-2035, July 2009.

[4.4] C.P.L. van Vroonhoven, D. D'Aquino and K.A.A. Makinwa, "A Thermal-Diffusivity-Based Temperature Sensor with an Untrimmed Inaccuracy of ±0.2ºC (3σ) from -55ºC to 125ºC", *IEEE ISSCC Dig. Tech. Papers*, pp. 314-315, Feb. 2010.

[4.5] C.P.L. van Vroonhoven and K.A.A. Makinwa, "Linearization of a Temperature Sensor Based on Thermal Diffusivity", *Proc. IEEE Sensors*, pp. 1697–1700, October 2009.

[4.6] S.H. Shalmany, "Thermal Diffusivity-Based Temperature Sensors for Microprocessors", M.Sc. Thesis, Delft University of Technology, Delft, The Netherlands, Oct. 2010.

[4.7] B. Razavi, "Design of Analog CMOS Integrated Circuits", McGraw-Hill, New York, 2001.

[4.8] C. P. L. van Vroonhoven, "Sigma-Delta Read-Out Architectures for Electrothermal Filters", M.Sc. Thesis, Delft University of Technology, May 2007.

[4.9] Feely, O.; Chua, L.O., "The effect of integrator leak in Σ-Δ modulation," *Circuits and Systems, IEEE Transactions on,* vol.38, no.11, pp.1293-1305, Nov. 1991

[4.10] C.P.L. van Vroonhoven and K.A.A. Makinwa, "An SOI thermal-diffusivity-based temperature sensor with ±0.6°C (3σ) untrimmed inaccuracy from −70°C to 225°C", *Sensors and Actuators A: Physical*, Volume 188, pp. 66-74, Dec 2012.

[4.11] C. P. L. van Vroonhoven, K. A. A. Makinwa, "A CMOS Temperature-to-Digital Converter with an Inaccuracy of ±0.5°C (3σ) from -55 to 125°C," *IEEE ISSCC Dig. Tech. Papers*, pp. 576 - 637, Feb. 2008.

[4.12] C. Zhang and K.A.A. Makinwa, "Interface Electronics for a CMOS Electrothermal Frequency-Locked-Loop", *IEEE J. Solid-State Circuits*, vol. 43, no. 7, pp. 1603–1608, July 2008.

5 Measurements

5.1 Introduction

This chapter presents measurement results for several single-ETF and multi-ETF temperature sensors. After discussing the measurement setup in section 5.2, section 5.3 will present measurements on silicon ETFs and attempt to validate the theory and modeling put forward in chapter 2. After that, section 5.4 shows the results for self-referenced sensors based on D_{Si} and D_{SiO2}.

The measurement data shown below were extracted from several test devices, developed over the course of this research. Three different technologies were available: two Bulk CMOS IC processes, with minimum feature sizes of 0.7µm and 0.18µm, and a 0.5µm SOI BiCMOS process. The bulk CMOS technologies were both single-well processes based on a p^+ substrate and a p^- epitaxial layer. The SOI technology was a partially-depleted BiCMOS SOI process with a minimum feature size of 0.5µm, and a thick buried oxide layer, located several microns below the surface.

Most devices were packaged in ceramic 16-pin DIL packages. These are easy to use, introduce little mechanical stress and work reliably over the extreme temperature range used in this work (-70ºC to 250ºC). To test for mechanical stress in plastic packages, 16-pin TSSOP packages were used.

Fig. 5.1 shows selected chip micrographs. The area of the chips was typically around $2mm^2$, limited by the area requirements of the padring. No substantial effort was made to optimize for die area. For the 0.18µm devices, the total active area was $0.18mm^2$; in the 0.7µm Bulk and 0.5µm SOI CMOS technologies, it was about $0.5mm^2$.

5.2 Measurement Setup

It is challenging to characterize temperature sensors (such as ETFs), because their temperature-sensing inaccuracy can only be accurately established by comparing their output with a more accurate *reference* temperature sensor that is in sufficiently close thermal equilibrium with the device under test (DUT). The paradigm of reducing all non-idealities to below a tenth of the lowest targeted inaccuracy (=±0.2 / 10 = ±0.02ºC) also holds for both the inaccuracy of the reference and the temperature difference between the reference and the DUT.

One way to ensure good thermal equilibrium is to include the reference sensor on the DUT, close to the ETF. However, the state-of-the-art inaccuracy of *integrated* temperature sensors, at ±0.1ºC (3σ) [5.1], is 5x too large; a more accurate reference sensor is thus required, and this is typically a discrete device. The reference temperature sensor used in this work is a Pt-100 thermistor that has been calibrated to better than 10mK inaccuracy (from –55ºC to 125ºC) at the Dutch Metrology Institute. By mounting the DUTs on a large aluminum block inside which the Pt100 was located, a sufficiently close thermal equilibrium was ensured.

The devices were characterized by sweeping temperature from -70°C to 170°C using a Vötsch VTM7004 climate chamber. At higher temperatures (125ºC – 250ºC), a Carbolite high-temperature oven was used. LabVIEW was used to automate all the measurements by controlling the climate chamber / oven in such a way as to ensure sub-0.02ºC thermal stability at several temperature points linearly spaced across the temperature range of interest. The lower temperature limit was set by the climate chamber, the upper limit by the robustness of the test PCB's dielectric layers.

In addition to oven-based testing, several devices were also tested in a Fluke temperature-controlled oil bath, in which temperature was swept from 20ºC to 150ºC. Such measurements are also very

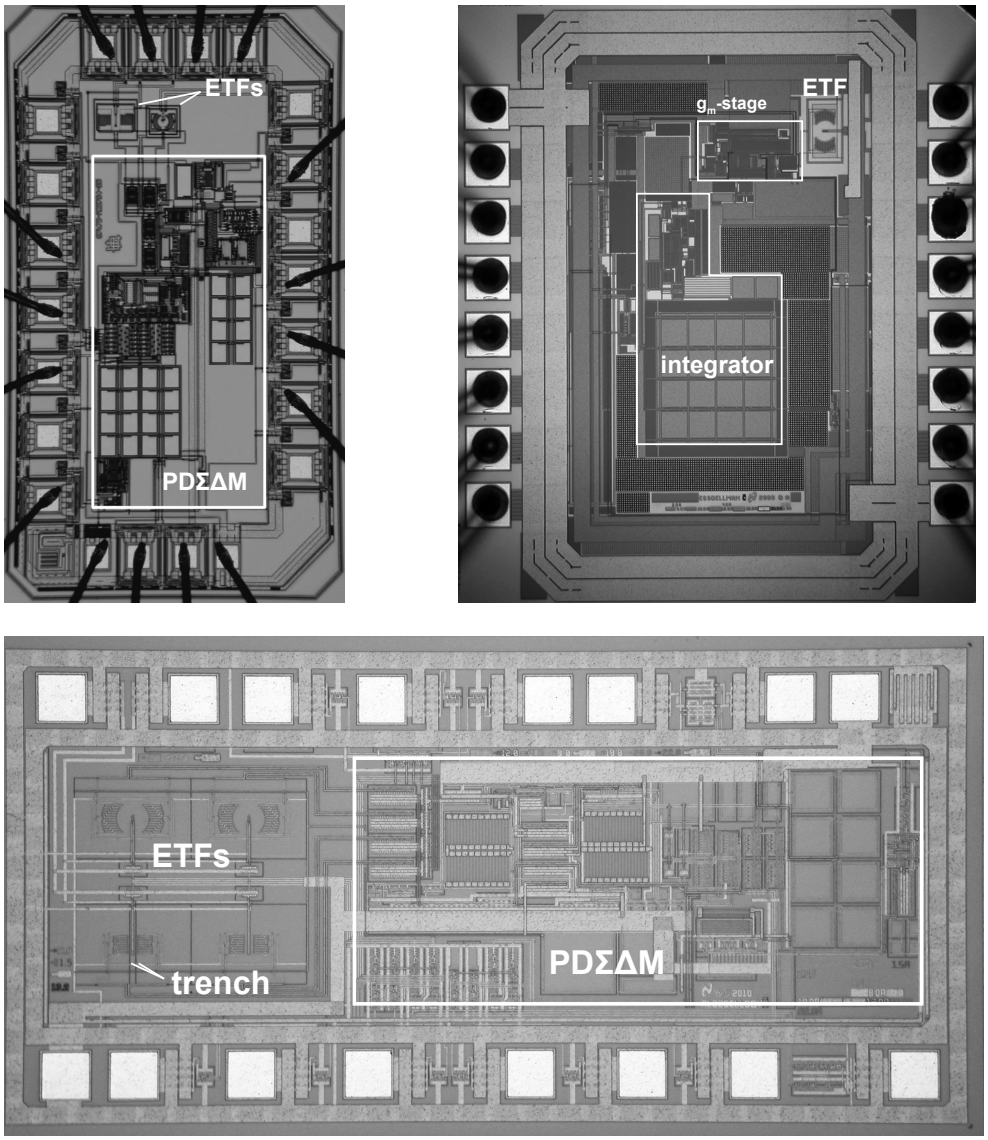


Fig. 5.1: Various chip micrographs. Top left: 0.7µm bulk CMOS; Top right: 0.18µm bulk CMOS; bottom: 0.5µm SOI CMOS.

accurate, since the DUTs are surrounded by constantly flowing oil that is kept at a well-defined temperature. Correlation measurements between the oven setup and the oil bath have shown errors at the ±0.05ºC level, indicating that both setups offer a high absolute accuracy.

When the measurement setup is thermally accurate, the only source of non-ETF temperature-sensing error is that of the heater drive and ETF readout circuitry. This has driven the precision circuit design discussed in Chapter 4, and Section 4.7 has already shown the low measured inaccuracy of a stand-alone PDΣΔM in 0.18µm CMOS. These measurements were done using the same PCB that was developed for testing the DUTs. For the other devices (in 0.7µm CMOS and 0.5µm SOI CMOS technology), a stand-alone PDΣΔM was not available, but (post-layout) simulation results show that

the phase error within one process corner, together with the residual offset of the PDΣΔM, are at the 0.05ºC level. Although there is incomplete coverage of heater drive and PDΣΔM performance for all devices, the Section 4.7 measurements and extensive measurements over P_{heat} and f_{drive} provide sufficient confidence to attribute most of the measured temperature errors to the ETFs.

5.3 Measurement results

5.3.1 Basic results in bulk and SOI CMOS technology

An important initial measurement is to establish ϕ_{ETF} as a function of temperature, and to see if ϕ_{ETF} does indeed have the expected $T^{0.9}$ temperature dependence, as suggested by D_{Si}. The main test vehicle used for this is the 'reference' ETF, which is a phase-contour ETF with s=24µm. Unless specified otherwise, the ETF is driven by f_{drive}=85.33kHz with P_{heat}=2.5mW.

Fig. 5.2 shows the measured ϕ_{ETF} for the reference ETF in 0.7µm bulk CMOS and 0.5µm SOI CMOS technology.

For bulk silicon, ϕ_{ETF} is about 20% higher than predicted by theory, suggesting that the effective D_{Si} is lower than originally predicted by the model. In related research, focused on modeling the ETF in greater detail [5.4], it was found that this is likely due to the lower D of the heavily doped substrate. Because the epi-layer thickness varies between process technologies, the effective D_{Si} can be expected to vary accordingly.

Despite a change in its nominal value, the temperature dependence of ϕ_{ETF} closely follows the predicted $T^{0.9}$ curve, as indicated by a linear fit on the double-logarithmic plot of $\phi_{ETF}(T)$ shown on the right hand side of Fig. 5.2. It should be noted that an exact fit should not be expected, since, as discussed in section 2.2.4, the $T^{0.9}$ "law" is itself only a convenient approximation of the actual temperature dependence of D_{Si}.

When implemented in an SOI process, ϕ_{ETF} clearly has a positive offset, associated with the increased thermal impedance introduced by the buried oxide layer. This corresponds to the behavior predicted in Chapter 2.

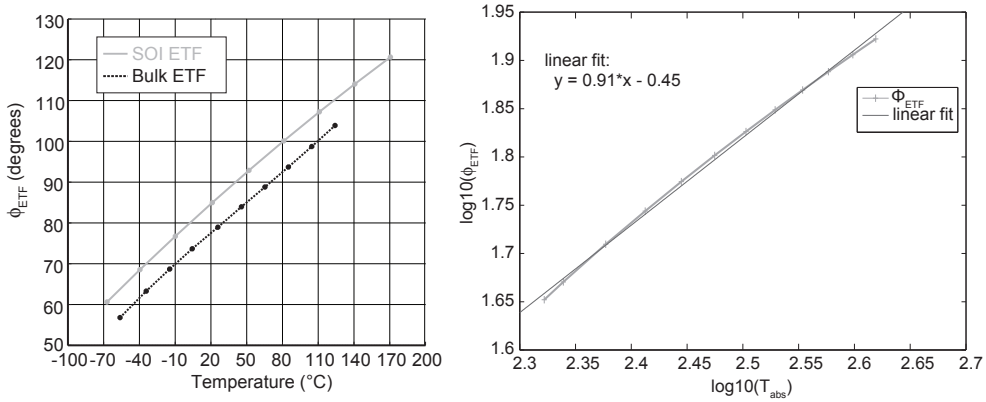


Fig. 5.2: Measured ϕ_{ETF} over temperature for an s=24µm ETF in bulk and SOI CMOS. Also shown is a double logarithmic plot showing the $T^{0.9}$ dependence of ϕ_{ETF} for a bulk silicon ETF.

In other measurements (not shown), the $T^{0.9}$ temperature dependence was also established for ETFs with smaller s. The proportionality of ϕ_{ETF} to f_{drive} was evaluated by sweeping f_{drive} while keeping temperature constant. Fig. 5.3 shows a good agreement with theory.

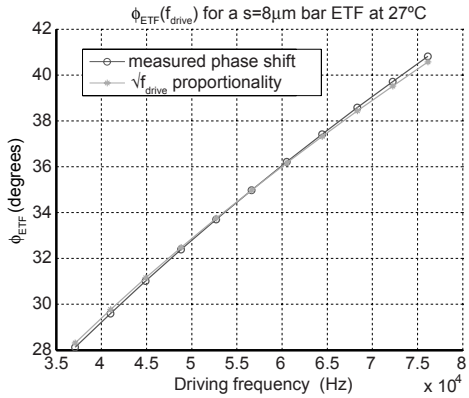


Fig. 5.3: Phase versus frequency behavior at constant temperature.

These initial measurements show that ϕ_{ETF} is a near-linear function of temperature, and validate the ETF's temperature sensing potential. Further measurements will focus on the accuracy of ETFs.

5.3.2 Effects of lithographic spread

The study of the effects of lithographic spread is one of the most important experiments performed in this work. In chapter 2, it was theorized that the dominant source of temperature error in ETFs should be due to the spread in *s* caused by random variations in lithography, and that ETFs should therefore become more accurate in more advanced, smaller feature size CMOS technologies.

The core experiment was to measure the untrimmed device-to-device spread for the phase-contour *s*=24μm silicon ETF in different process technologies. To calculate temperature from ϕ_{ETF}, the measured digital output was applied to a 5th-order polynomial fit in MATLAB. To obtain device-to-device spread, the digital output of each of these devices was then fed to the polynomial, and the resulting temperature output was compared to the reference temperature measurement.

The measurement results are shown below in Fig. 5.4. For both technologies, 16 devices were measured. The thick black lines indicate the estimated ±3σ error limits based on this dataset.

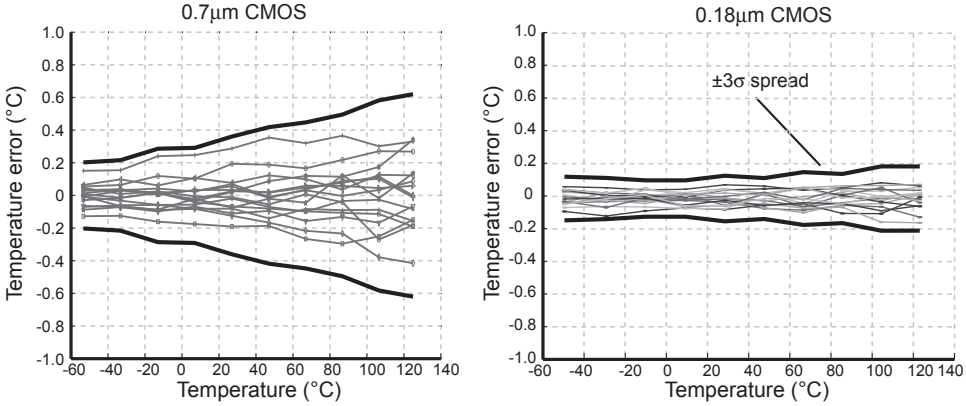


Fig. 5.4: Measured untrimmed device-to-device spread for the *s*=24μm ETF.

The above results show that lithography is indeed the main source of temperature in ETFs and that in advanced processes, an untrimmed inaccuracy as low as ±0.2ºC (3σ) can be achieved over the over

the -55ºC to 125ºC range. This compares favorably to the state of the art in batch-calibrated BJT-based temperature sensors.

For the same ETF in the 0.5µm SOI process, the results are similar (Fig. 5.5 A). These results are also based on measurement data from 16 devices. The data of these ETFs can be combined in a plot showing inaccuracy (at 125ºC) versus technology node for two different s (Fig. 5.5 B). Where available, data for a scaled ETF with $s \approx 12\mu m$ was also added.

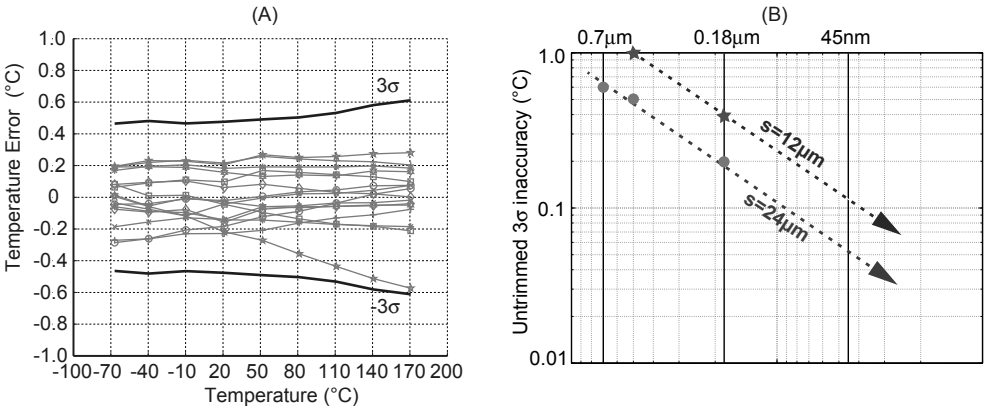


Fig. 5.5: (A) Device-to-device spread in SOI and (B) plot showing inaccuracy versus technology node.

Extrapolating the trend shown above, ETFs benefit greatly from technology scaling and fine line geometries. However, in scaled technologies, s=24µm ETFs will quickly achieve inaccuracies that are very hard to measure and, perhaps more importantly, are not typically required for integrated temperature sensors. In that case, rather than keeping s constant, ETFs in deep submicron CMOS should have a strongly scaled s, which improves their SNR.

5.3.3 Resolution and power consumption

Although ETFs are very accurate, their power efficiency is fundamentally limited, since the sensors have to heat the silicon die to measure temperature. Chapter 2 has presented how resolution is a function of s, and how the use of a buried oxide layer may improve power efficiency by blocking some of the heat that would otherwise simply get lost to the substrate.

This section will study the power efficiency of ETFs in different technologies and of different geometry. The power consumption of the readout circuit is not taken into account, in order to extract a more clear result of the scaling properties and the technology dependence of the ETF itself. This can be justified since on all test chips, the ETF was thermal noise-limited by its thermopile, and not by its readout circuit.

It is useful to define a figure of merit (FOM) to facilitate a straightforward comparison. The standard resolution FOM for temperature sensors is given by [5.6]:

$$FOM1 = \frac{Energy}{conversion} resolution^2 \tag{4.19}$$

Here, energy per conversion is power consumption multiplied by conversion time. A lower FOM indicates higher performance. Although FOM1 works well for comparing ETF-based to other temperature sensors [5.6], it is not as suitable for comparing ETFs between each other directly. This is because in ETFs, P_{heat} is linearly proportional to V_{ETF} and thus to the inverse of resolution. For a

doubling of P_{heat}, the FOM1 halves, while it is preferable to use a FOM that is P_{heat}-independent, since such a FOM would more clearly capture the intrinsic performance of the ETF.

A more suitable FOM can be defined as follows:

$$FOM2 = \frac{resolution * P_{heat}}{\sqrt{f_{conv}}}, \tag{4.20}$$

in which f_{conv} is the conversion rate, in decimated samples per second, resolution is the noise in ºC, and P_{heat} is the heater power. (A lower FOM2 corresponds to better performance.) Table 5.1 shows the FOM2 of several ETFs, with different s and in different technologies:

Table 5.1: Measured resolution for different ETFs.

	ETF1	ETF2	ETF3	ETF4	ETF5	ETF6
Tech. node (µm)	0.7	0.18	0.18	0.7	0.5 SOI	0.5 SOI
distance *s* (µm)	24	24	12	10.5	24	24
f_{drive} (kHz)	85	43	85	341	85	85
Conversion rate (S/s)	0.16	0.16	0.16	1.30	0.65	0.65
Power (mW)	2.50	2.50	0.83	2.50	1.00	1.00
Noise (ºC$_{rms}$)	0.03	0.02	0.03	0.026	0.03	0.021
FOM2 (mK*mW/√Hz)	186	124	62	57	37	26

ETF1 is the reference ETF in 0.7µm CMOS, which has the highest FOM. ETF2 has the same s but was made in 0.18µm CMOS; this technology has a lower R_{tp}, and requires less thermocouple spacing to accommodate DRC rules, which reduces the thermopile's thermal noise. ETF3 is a scaled ETF, which has significantly better FOM. A similar result was achieved for an aggressively scaled ETF in 0.7µm CMOS, ETF4. This ETF was driven at 341kHz and achieves about 3.5x better resolution than the original ETF1.

The s=24µm ETF in SOI has a 5x better FOM than the same ETF in bulk CMOS. ETF6 is an optimized ETF with 2x more thermocouples and represents a further 1.4x improvement. This 7x SNR improvement demonstrates the feasibility of wide-range thermal diffusivity sensors with sub-mW power dissipation in SOI technology.

ETF theory predicts that resolution is a function of both s and f_{drive}, and that a weak optimum exists for each geometry. Fig. 5.6 compares the measured resolution of s=24µm and s=12µm bulk CMOS ETFs as a function of f_{drive}. To measure noise more easily, P_{heat} was only 1mW, and the conversion rate was set to 1S/s.

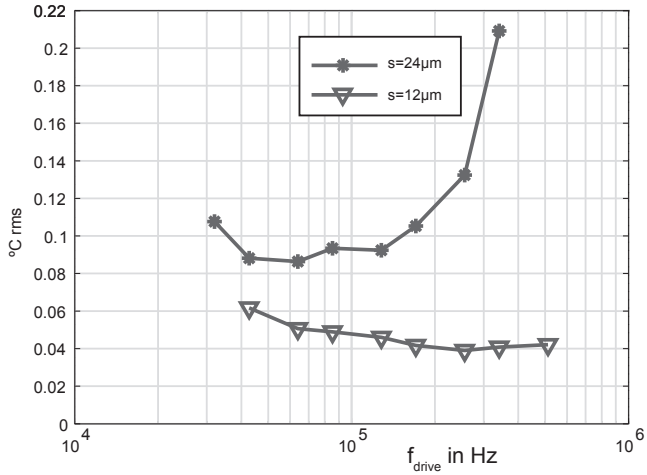


Fig. 5.6: Measured resolution as a function of f_{drive}, for two ETFs.

As can be seen, the scaled ETF offers increased resolution, and the optimal f_{drive} is at a higher frequency than for the reference ETF. At low frequencies, low $d\phi_{ETF}/dT$ and residual low-frequency noise degrade the resolution. For the reference ETF, operating at very high f_{drive} reduces the amplitude of V_{ETF} significantly. These results agree well with simulations (see Fig. 2.38) .For further measurements of the resolution of scaled ETFs (with small s), the reader is referred to [5.5].

5.3.4 Sensitivity to doping fluctuations

It is known that the thermal conductivity of silicon is a much weaker function of doping than is its electrical conductivity, and it was argued that this is one of the main attractive features of ETFs.

In prior art [5.2], it was shown that the thermal diffusivities of p-doped silicon ($N_a = 5*10^{14}/cm^3$) and n^+-doped silicon ($N_d = 3*10^{16}/cm^3$) are equal to within 0.1%. This was established by measuring 16 devices over temperature, each having two s=20µm bar ETFs implemented: one in which the heat diffused through a lightly doped p-substrate, the other in which the heat diffuses through an n-well. The n-well ETF was measured to have an offset of only 0.2ºC relative to that of the substrate ETF, showing, indeed, a low sensitivity to doping fluctuations. At these relatively low doping levels, however, it is difficult to accurately measure associated changes in thermal conductivity and diffusivity at 'high' (T>200K) temperatures [5.3].

In this work, a second experiment was performed using an n^{++}-implant (Fig. 5.7) with a doping concentration that is more than 2 orders of magnitude higher than in [5.2]. This implant was available in the form of the buried collector implant of the SOI BiCMOS process.

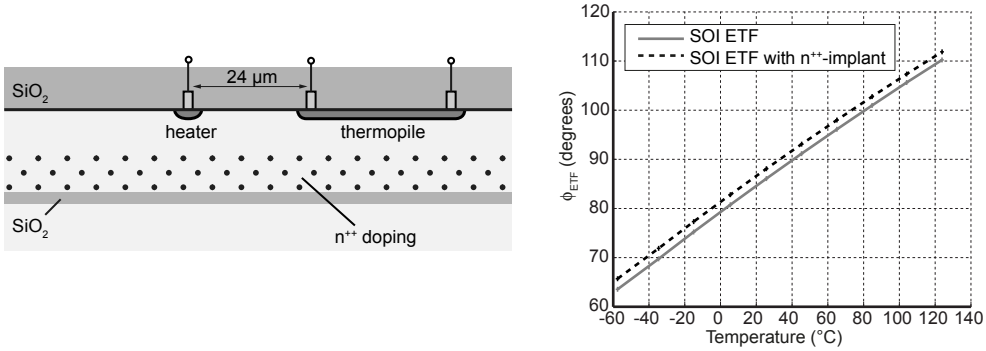


Fig. 5.7: Illustration of n^{++}-doping ETF and measured associated increase in ϕ_{ETF}.

This high doping level leads to a measureable increase in ϕ_{ETF} of ~2.3 degrees, corresponding to ~9°C (Fig. 5.7). In contrast to the regular ETF, this ETF was found to have significant batch-to-batch errors of up to 0.8°C shift in average, as shown in Fig. 5.8. This shift can clearly be attributed to spread in doping concentration, since all other parameters are identical.

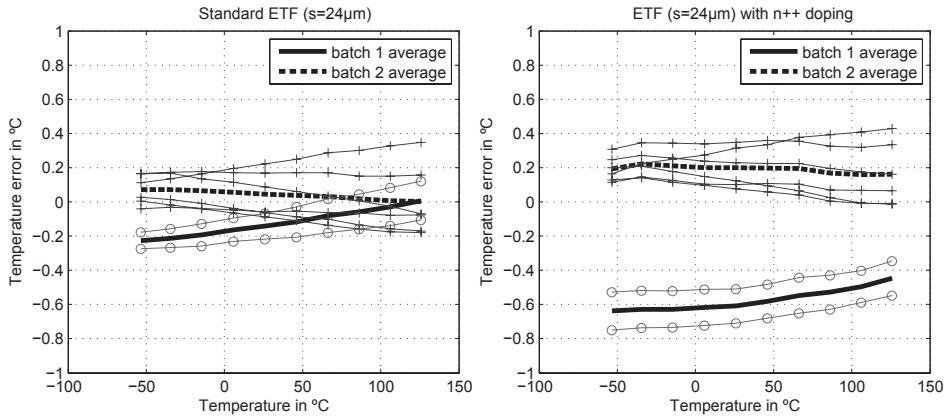


Fig. 5.8: Batch-to-batch spread for an undoped and heavily doped s=24µm ETF in SOI CMOS.

A small CTAT batch-to-batch component is common for both the standard and the n^{++} ETF. From theory, this is neither doping nor lithography-related, and possibly points to an effect associated with the buried layer thickness and/or depth.

The temperature offset introduced by spread in doping corresponds approximately to the predictions made in section 2.5.6. Since typical doping concentrations are more than 2 orders of magnitude lower, this reaffirms that for regular ETFs, the effect of doping is small.

5.3.5 Wafer-to-wafer and batch-to-batch spread

To evaluate the effects of wafer-to-wafer spread, 24 devices from two separate wafers, fabricated in a 0.18µm Bulk CMOS process, were measured over temperature. Fig. 5.9 shows the results, indicating that the averages are no more than 0.1ºC apart over the military temperature range.

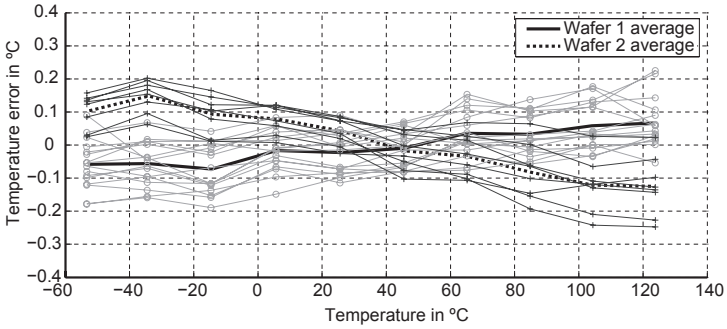


Fig. 5.9: Wafer-to-wafer spread for the s=24µm ETF in 0.18µm bulk CMOS.

To evaluate the effect of batch-to-batch spread, the *s*=24µm ETF was fabricated on two separate multi-project-wafer (MPW) runs, the second 8 months after the first. The readout circuitry was identical for both devices.

Rather surprisingly, since lithography was thought to be the dominant source of error, the *s*=24µm ETF in 0.18µm bulk CMOS showed significant batch-to batch spread. As shown in Fig. 5.10, the devices exhibited a batch-to-batch offset of about 2°C, while the intra-batch spread remained at ±0.2°C (3σ).

The same experiment was done for the *s*=24µm ETF in the SOI process. Fig. 5.10 (B) shows the measured results for 32 devices from 3 process lots, manufactured over a 1-year period. Interestingly, the batch-to-batch variations are much smaller, and a worst case shift in average of about 0.2°C was observed. These ETFs, therefore, are truly *untrimmed* temperature sensors with an inaccuracy better than ±1.0ºC (3σ).

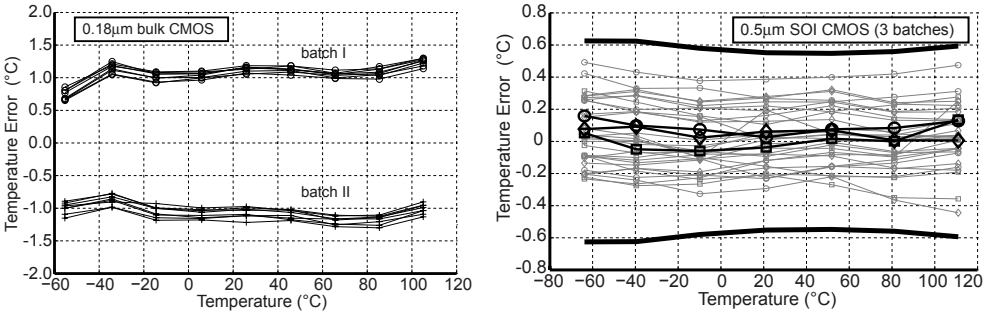


Fig. 5.10: Measured batch-to-batch spread for 0.18µm Bulk CMOS (2 batches) and the 0.5µm SOI CMOS process (3 batches).

The most likely cause for the observed batch-to-batch offset in bulk CMOS is a combination of spread in the thickness of the lightly-doped epi-layer and spread in the doping level of the heavily doped p^+-substrate. It is known that the thermal conductivity of heavily doped silicon is about 20–30% lower than that of lightly doped silicon [5.4]. Since the epi-layer is typically only a few microns thick while the spacing between the heater and the thermopile is as much as 24µm, the ETF may be sensitive to both the thickness of the epi-layer and the doping level of the substrate [5.4]. This hypothesis is strengthened by the measurements in SOI CMOS, since the presence of the thermally insulating buried oxide ensures that heat diffusion in the ETF is dominated by the well-defined diffusivity of the lightly-doped epi-layer.

By applying a batch offset trim to the bulk CMOS devices, the residual batch-to-batch error can again be reduced to ±0.2°C over a –15°C to 125°C range. At lower temperatures, this error increases to ±0.6°C. This is shown below:

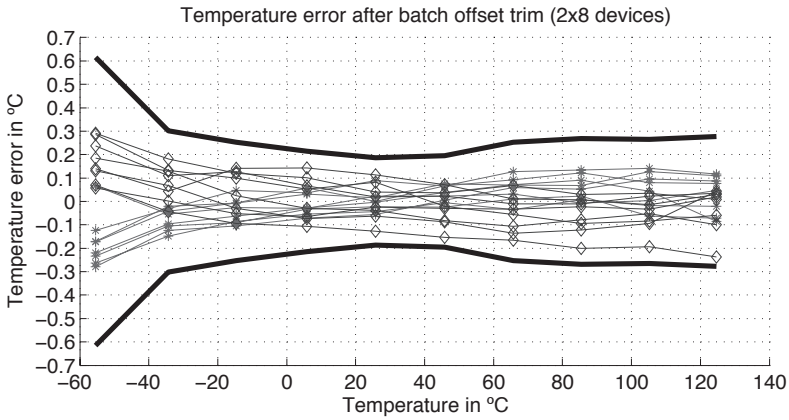


Fig. 5.11: Device-to-device spread after a batch offset trim at room temperature (0.18µm bulk CMOS, 2 batches).

Although these ETFs still require trimming to compensate for batch-to-batch variations, this can be achieved by relatively simple (digital) offset trim at room temperature. Furthermore, the value of the offset can be established by a batch-calibration, i.e. by measuring a few devices at a well-known temperature. The offset trim does not effectively reduce the curvature error at lower temperatures, but in thermal management applications, low-temperature errors are typically less important.

It would be of interest to study batch-to-batch variations for ETFs with smaller s, since these should be less sensitive to variations in the epi-layer thickness. In this work, insufficient data were available to further study this.

5.3.6 Mechanical stress

To test for the effects of mechanical stress, devices from a single wafer were packaged in both 16-pin ceramic DIL and 16-pin plastic packages. The plastic packages were fabricated using the "scoop and goop" method: first, empty (no-die) 16-pin TSSOP packages were formed by surrounding a leadframe with mold compound. These packages were then etched open to re-expose the leadframe and the leadfingers. The dies were then placed onto the leadframe and the pads bonded to the leadfingers. Finally, a new mold compound was injected to seal the package again. As the package cools down, the mold compound contracts and exerts compressive stress onto the die. Although this is only a qualitative test, it is a useful way to screen out any serious issues associated with mechanical stress.

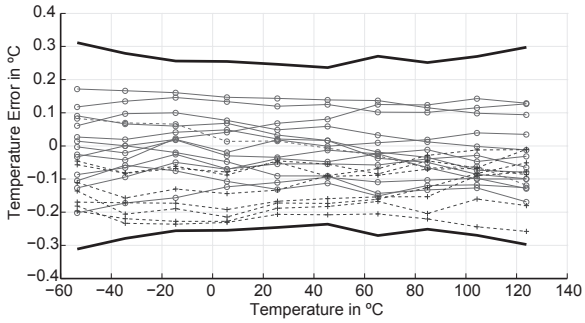


Fig. 5.12: Measured device-to-device errors for devices in plastic packages.

The above figure shows the device-to-device temperature errors of 16 devices in plastic packages (straight line) and 8 devices in ceramic packages (dashed lines). All 24 devices came from the same wafer, so that wafer-to-wafer spread was not present. The 3σ limits are based on the 16 devices in plastic packages, and these show a slightly increased spread of ±0.3ºC (3σ) as compared to the same ETF in ceramic packages (±0.2ºC (3σ), see Fig. 5.5). It can also be seen that the 8 non-plastic devices have a mean negative temperature error, indicating that they underestimate the ambient temperature. From these two observations, the most likely reason for the slight increase in spread is an increased spread in self-heating. From the mean temperature error and given that P_{heat} =2.5mW, the R_{th} of the plastic packages is about 80K/W higher than for the ceramic packages. The absolute spread in R_{th} is likely higher than for the ceramic packages, thus increasing the device-to-device spread slightly. It is unlikely that the ETFs themselves have a stress-induced offset in ϕ_{ETF}, but even if this were the case, the shift is at the 0.1ºC level.

Although this is only a qualitative experiment, first results show that ETFs are essentially stress-insensitive, as predicted by theory (see section 2.5.2.8). This implies that trimming (if required) can be done at wafer level, which reduces cost significantly.

5.3.7 Self-Heating

As already suggested in the previous section, for high P_{heat} or high R_{th}, the effects of spread in self-heating can be a significant source of error. This is exacerbated in SOI technology, since the buried oxide layer locally increases the R_{th} seen by the ETF.

Chapter 2 has predicted that ϕ_{ETF} will be determined by the effective D_{Si} along the thermal path, which has to be evaluated as a function of the DC thermal gradient between heater and sensor. Since the thermal gradient is non-linear, sensing closer to the heater increases the ϕ_{ETF} dependence on P_{heat}. This can be expressed as a finite R_{th}, i.e. a temperature error induced by P_{heat}.

Fig. 5.13 shows the measured R_{th} for bar ETFs in SOI with 4 different values for s. To establish R_{th}, P_{heat} was swept from 2.5mW to 15mW, and the relative change in digital output (mapped to temperature) was used to calculate R_{th}. As suggested by theory, ETFs with smaller s sense an increased R_{th}. Fortunately, the R_{th} only approximately doubles when reducing s by 3x. This means that for scaled ETFs, the resolution/self-heating tradeoff actually becomes less stringent.

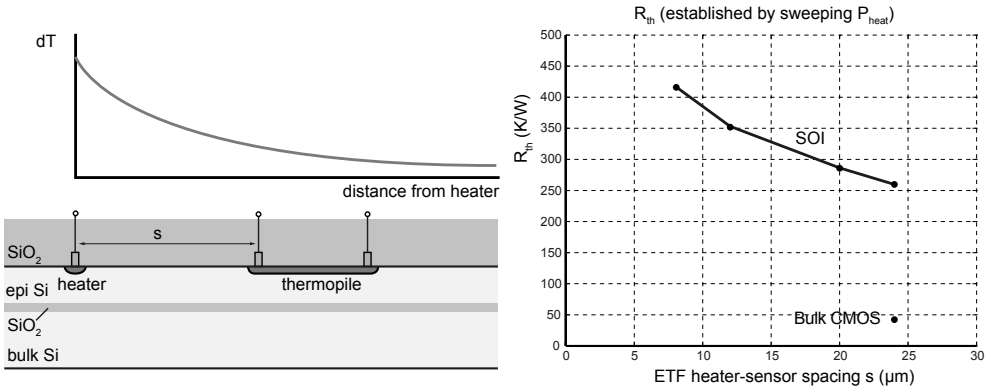


Fig. 5.13: Self-heating measurements for ETFs in SOI. The plot shows the effective R_{th} as a function of s.

Local self-heating increases the ETF's sensitivity to P_{heat}. When P_{heat} is generated by switching V_{DD}, this introduces a finite power supply rejection (PSR). Measurements indicate that a spread of 1mW for an s=8µm ETF in SOI leads to a 0.4ºC error. Since P_{heat} is proportional to V_{DD}^2, PSR degrades at higher V_{DD}. One way to reduce this sensitivity is to use an LDO or a P_{heat} control circuit, but these are typically not power-efficient. In future work, V_{DD}-dependent duty cycling of the ETF may improve PSR.

For temperature sensors that measure ambient temperature, self-heating may limit the sensor's inaccuracy and place an upper bound on P_{heat} and/or a minimum bound on s. For ETFs that sense die temperature (as in thermal management applications), self-heating is less critical, as long as the circuitry of which the temperature is to be sensed is in good thermal contact with the ETF.

5.3.8 Temperature range

Since the temperature information provided by ETFs is in the time domain, they are fundamentally insensitive to leakage, and thus should be able to operate at very high temperatures.

The bulk CMOS test devices were mostly used to show the effects of lithography and as such, were not optimized for high-temperature performance. However, as Fig. 5.14 shows, a reference ETF in 0.7µm CMOS operates over the -70 to 160ºC range, and still achieves very good untrimmed inaccuracy. Even at very high or very low temperatures, ϕ_{ETF} still follows its D_{Si} temperature dependence, and the device-to-device spread is still well-defined.

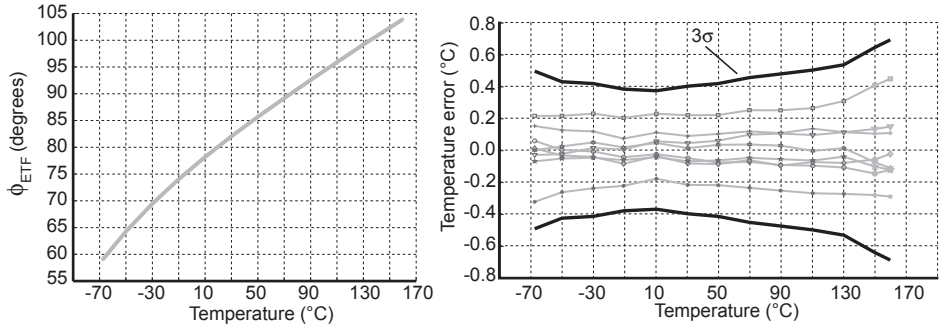


Fig. 5.14: ϕ_{ETF}(T) and wide-range results for an s=20µm bar ETF in 0.7µm CMOS (8 devices).

The bulk CMOS devices were not tested extensively at extreme temperatures (>160ºC). The SOI devices, on the other hand, were measured at temperatures up to 225ºC. The resulting untrimmed temperature error for a bar ETF structure with s=20µm is shown below:

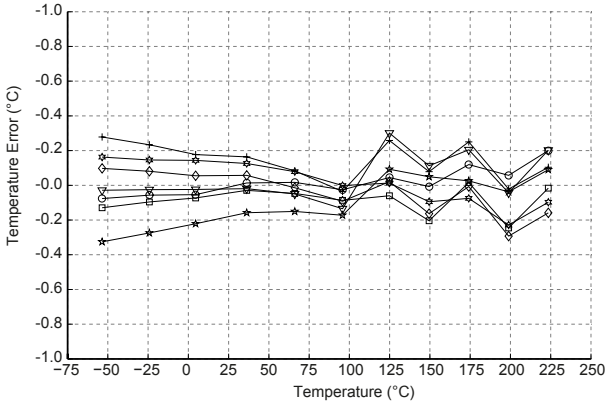


Fig. 5.15: Device-to-device errors for an s=20µm bar ETF in SOI, over the -55 to 225ºC range (7 devices).

To span this wide temperature range, two different temperature-controlled climate chambers were required. The somewhat erratic behavior at very high T is attributed to the temperature stability of the high-temperature oven, which was significantly lower than for the regular temperature range setup.

The above results show that ETFs are excellent wide-range temperature sensors. Their operating range was limited by the measurement setup in all cases, and so their operating range can likely be extended below -70ºC and above 225ºC. In one demonstration experiment, ETFs were operated at 260ºC. Even at these high temperatures, ETFs should still maintain their excellent untrimmed inaccuracy, making them attractive for high-temperature industrial applications.

5.3.9 Thermal Interference

When used in ASICs, other heat sources on the die may potentially interfere with the temperature variations detected by the ETF's thermopile. Because the readout circuit is based on synchronous detection, this is only a problem if the heat is generated at f_{drive} (or its odd multiples), and only if the source of interference is close to the thermopiles, since its heat output would otherwise be thermally low-pass filtered by the silicon substrate.

The ETF's sensitivity to interference was tested by driving a neighboring heater located 120µm away with a 12mW pseudo-random heat signal derived from f_{drive}; this test is illustrated in Fig. 5.16:

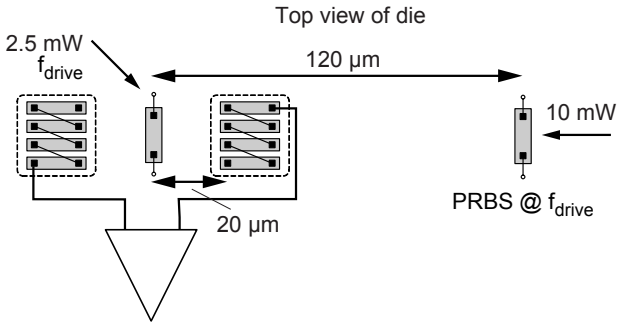


Fig. 5.16: Driving a parasitic heat source close to an ETF to test for thermal interference.

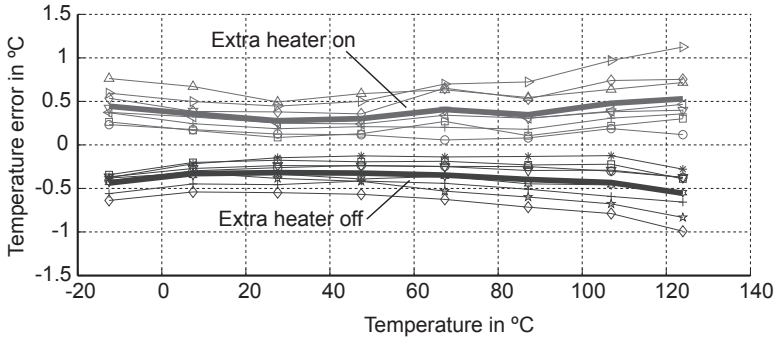


Fig. 5.17: Measured temperature error in the presence of a large parasitic heat source.

Fig. 5.17 shows the results. The device spread is relatively large, since this test was done for 0.7µm test devices. While an increased self-heating of about 0.5°C was measured, the ETF's resolution or its inaccuracy did not degrade significantly, demonstrating that most of the thermal interference was indeed filtered by the thermal inertia of the die. The increase in die temperature is due to an effective R_{th} of about 50K/W, and this data point is also shown in Fig. 5.13. The absence of a buried oxide layer and the high thermal conductivity of the ceramic package lead to this relatively low R_{th}.

5.4 Ratiometric ETF-based temperature sensors

As outlined in chapter 3, a self-referenced ETF-based temperature sensor performs a ratiometric measurement of D_{Si} and D_{SiO2} to measure absolute temperature *without* requiring an accurate (external) time reference. With silicon ETFs being well-characterized, the performance of such sensors depends strongly on the behavior of oxide ETFs. Preliminary device characterization showed that ETFs in which ϕ_{ox} is measured via the SiO_2 in the sidewalls of SOI deep isolation trenches tend to have the most well-defined ϕ_{ox}, and therefore these were studied further.

5.4.1 Basic results

The idea of using a deep trench isolation structure to obtain D_{SiO2} has been studied in several 0.5µm SOI devices. Two bar ETFs were implemented, with one having a ≈1µm trench in between heater and sensor. The heater-sensor spacing was set to the minimum allowed by DRC rules, which was s=8µm. The devices were measured over a wide temperature range (-70ºC to 200ºC), limited by the measurement setup.

Fig. 5.18 shows the measured phase characteristics of both ETFs and their difference, $\phi_{ETF1} - \phi_{ETF2} = \phi_{trench}$, when driven at $f_{drive} \approx 50$kHz.

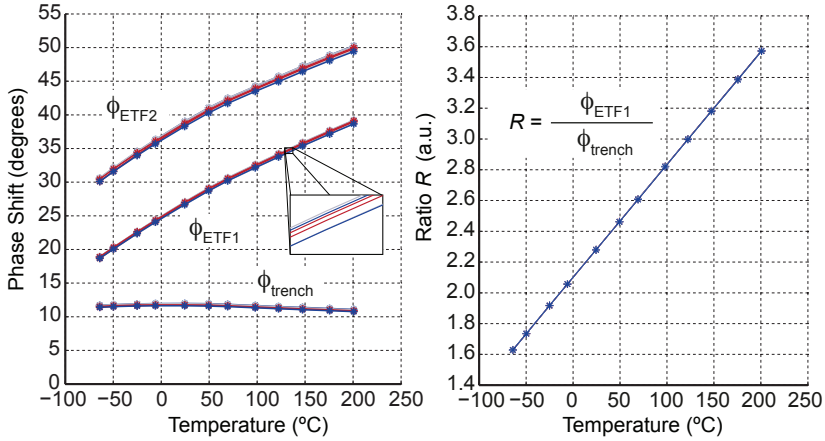


Fig. 5.18: Phase measurements on s=8µm ETFs.

Both ETFs can be seen to have a very similar $\sim T^{0.9}$ temperature dependence. As expected, the oxide sidewalls of the trench cause an additional phase shift in ϕ_{ETF} that is nearly temperature independent; its value is in line with the expected sidewall thickness.

5.4.2 Frequency sensitivity

Since each ETF is driven by an inaccurate on-chip RC oscillator, the device-to-device spread on ϕ_{ETF1} or ϕ_{ETF2} is large and poorly defined. However, as predicted, the f_{drive} sensitivity of the *ratio* of two ϕ_{ETF}s is much less sensitive to frequency. The right hand side of Fig. 5.18 plots the ratio of ϕ_{ETF1} and ϕ_{trench}, and it can be seen that this is a near-linear function of temperature. In order to understand how much this sensitivity is reduced, both the ϕ_{ETF1} and the $R=\phi_{ETF1}/\phi_{trench}$ curves were fitted to temperature at f_{drive} = 50kHz, after which f_{drive} was swept from 37 to 77kHz. The resulting temperature errors are shown in Fig. 5.19:

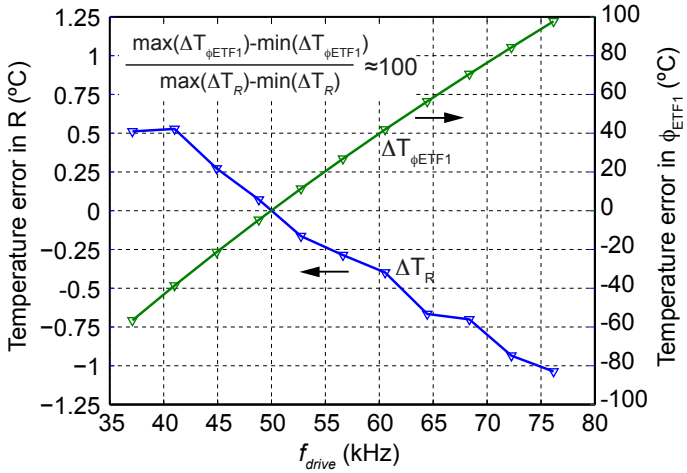


Fig. 5.19: Frequency sensitivity of ϕ_{ETF1} and R.

The spread of the on-chip oscillator, at ±2.5% (min – max, based on 12 devices), translates into a ±5ºC error for ϕ_{ETF1}, but only a ±0.1ºC error in R, showing that the ratiometric approach reduces f_{drive} sensitivity by about 50x at f_{drive}= 50kHz, while it is about 100x over the full f_{drive} test range. The efficacy of this method is limited by the fact the Si and SiO_2 heat diffusion paths deviate slightly from the ideal $\phi_{ETF} \propto \sqrt{f_{drive}}$ proportionality.

5.4.3 Temperature error

Temperature error and device-to-device spread were established by comparing the output of each device to a master curve. The master curve was obtained by fitting all R data from all devices to the temperature data from the reference sensor. Having obtained this master curve, the device-to-device spread was now computed as the error of each device to the master curve. The untrimmed device-to-device spread is shown in Fig. 5.20:

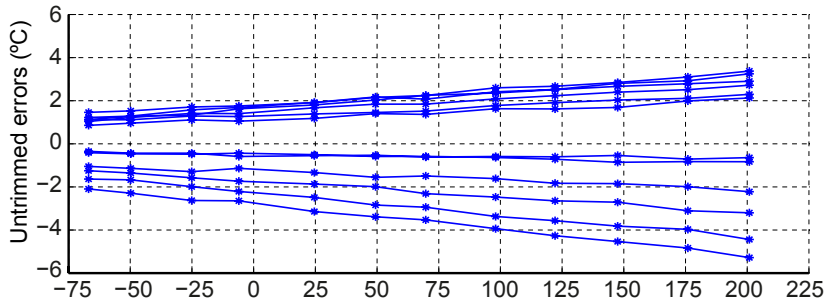


Fig. 5.20: Untrimmed device-to-device temperature errors in *R*.

One of the main reasons for these relatively large untrimmed errors is the spread in the thickness of the deep trench isolation sidewalls. However, this is also the dominant source of error and the main degree of freedom in R, and so the temperature inaccuracy can be greatly improved by applying a single trim. Since spread in t_{ox} causes an offset in ϕ_{ox} and therefore a gain error in R, a gain trim is most effective (as also indicated by the Fig. 5.20 results).

Fig. 5.21 shows that a single-temperature digital gain trim at 25ºC reduces the spread to ±0.4ºC (3σ) over the -70ºC to 200ºC temperature range. These results are based on 12 devices from a single process lot.

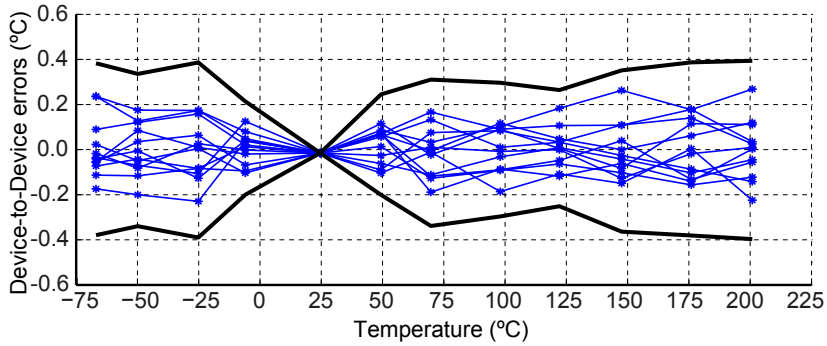


Fig. 5.21: Device-to-device temperature errors after a single-temperature gain trim.

Since the above errors are calculated with respect to a master curve for R, they do not take into account nonlinearity in R. However, as discussed in chapter 3, the self-referenced system can be tuned for optimal linearity. A linear fit to the measurement data leads to the following error plot:

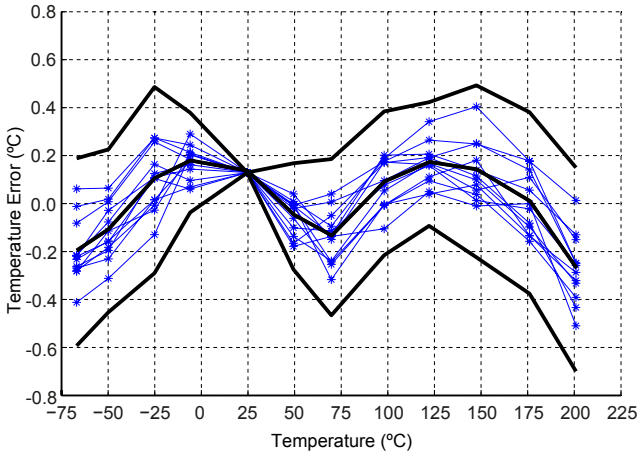


Fig. 5.22: Temperature error to a linear fit; the thick lines indicate mean and ±3σ limits.

Fig. 5.22 shows that the nonlinear error is within ±0.25ºC over the full 270ºC temperature range. A more linear output can be obtained using the analog and/or digital linearization / curvature correction techniques described earlier in this thesis. A linear fit optimized for a smaller temperature range will also have reduced error (e.g. ±0.1ºC from -40ºC to 105ºC).

5.4.4 Resolution

The resolution of this sensor is limited by thermal noise from the thermopiles and front-end of both ETF1 and ETF2, and by how these noise sources affect R. Section 3.5.4 has shown how errors in ϕ_{ox} (ϕ_{trench} in this case) are amplified by the ϕ_{Si}/ϕ_{ox} ratio: noise in ϕ_{trench} therefore dominates the total noise. When both ETF1 and ETF2 are continuously operated at P_{heat} = 2.5mW, the temperature-sensing resolution was measured to be 0.075ºC$_{rms}$ at 0.7 conversions/s. This is about 10x noisier than the best single-ETF sensors, while being orders of magnitude worse than the resolution of BJT-based temperature sensors. The limited resolution of ETF-based temperature sensors is an important field for further research.

5.5 References

[5.1] M.A.P. Pertijs, K.A.A. Makinwa and J.H. Huijsing, "A CMOS temperature sensor with a 3σ inaccuracy of ±0.1ºC from -55ºC to 125ºC", *IEEE J. Solid-State Circuits*, pp. 2805-2815, Dec. 2005.

[5.2] C. Zhang and K.A.A. Makinwa, "The effect of substrate doping on the behaviour of a CMOS electrothermal frequency-locked-loop", Dig. Transducers 2007, pp. 2283-2286, Jun. 2007.

[5.3] M. Asheghi et al., "Thermal conduction in doped single-crystal silicon films", Journal of Applied Physics, Vol. 91, pp. 5079-5088, April 2002.

[5.4] B. Vermeersch, "Thermal AC Modelling, Simulation and Experimental Analysis of Microelectronic Structures Including Nanoscale and High-Speed Effects," PhD dissertation, University of Gent, 2009.

[5.5] S.H. Shalmany, "Thermal Diffusivity-Based Temperature Sensors for Microprocessors", M.Sc. Thesis, Delft University of Technology, Delft, The Netherlands, Oct. 2010.

[5.6] K.A.A. Makinwa, "Smart Temperature Sensor Survey", [Online]. Available: http://ei.ewi.tudelft.nl/docs/TSensor_survey.xls

6 Conclusions and Outlook

This chapter will list the main findings and most important conclusions of this thesis. It also provides some suggestions for future work.

6.1 Main findings / Original Contributions

The main findings of this thesis are summarized below. Unless indicated otherwise, these findings are original contributions that expand the body of knowledge on Electrothermal Filters (ETFs) and define the state of the art.

6.1.1 Electrothermal filters

- ETFs can be used to accurately measure the thermal diffusivity of silicon and silicon dioxide. With regard to measuring D_{Si}, this thesis expands on the prior art with improved modeling and more accurate measurement results from different process technologies, and for several ETF structures.
- With regard to measuring D_{SiO2}, this thesis presents the first results towards using D_{SiO2} for temperature sensing. Two types of oxide ETFs were presented: a first type measures the diffusion of heat through the field oxide / shallow trench isolation between a polysilicon layer and the substrate; a second type is based on the heat diffusion through the oxide sidewalls of deep trench isolation structures in SOI technology. Both show that the thermal delay of oxide ETFs is nearly temperature independent. This enables the realization of self-referenced thermal diffusivity temperature sensor, but also opens up possibilities for thermal-domain time references (the "thermal bandgap").
- Semi-analytical ETF models were developed which provide a better understanding of the precision and resolution of ETFs, together with estimates for their scaling behavior, also including secondary effects such as self-heating.
- The device-to-device spread of silicon ETFs has been studied for different geometries and process technologies. It was found that the lithographic inaccuracy with which ETFs are fabricated is the dominant source of error, and thus that a near-proportional relationship between inaccuracy and CMOS technology process node exists.
- The resolution of ETFs and its dependence on ETF size and geometry have been studied, a more-than-linear scaling behavior with regard to improved resolution for smaller ETFs was found. A highly sensitivity relative temperature sensor and a low-noise readout circuit remain critical for high resolution / conversion rates.
- The operating range of ETFs has been extended significantly, down to -70ºC and up to as high as 250ºC. This shows that ETFs are a good candidate for fully integrated extreme temperature sensors, which may find applications in e.g. downhole drilling and other industrial applications.
- Measurement results on the effect of doping expand the evidence that ETFs are relatively insensitive to doping fluctuations, but also that ETFs with very high substrate doping levels (10^{19} atoms/cm^3) have significant batch-to-batch spread.
- Qualitative measurement results show that the effect of mechanical stress is limited and that no "packaging shift" is observed.
- This work presents the first accurate data on the performance of (scaled) ETFs in SOI technology; results show device-to-device spread, doping insensitivity, negligible batch-to-batch spread, greatly improved temperature sensing resolution and very wide operating range (up to 250ºC).
- For silicon ETF sensors that have a time reference available, an untrimmed temperature-sensing inaccuracy of ±0.2ºC (3σ) over the -55ºC to 125ºC range was measured.
- For self-referenced sensors based on both D_{Si} and D_{SiO2}, an inaccuracy of ±0.4ºC (3σ) over the -70ºC to 200ºC range was measured; a single trim at room temperature was required.

6.1.2 Readout circuitry

- This thesis has shown that the phase shift of the small ETF output signal can be digitized with high precision and high resolution by using frequency- and phase-domain sigma-delta modulators and standard decimation filters. These ΣΔMs expand the class of time-to-digital converters based on sigma-delta modulation.
- Phase-domain ΣΔ modulators offer better linearity than their frequency-domain counterpart, and are thus preferred for ETF readout.
- As with all ΣΔ modulators, the oversampling nature of the PDΣΔM and the presence of a decimation filter with well-defined notches combine favorably with dynamic offset cancellation techniques such as chopping, enabling a low input-referred offset while not affecting the input noise density.
- For small sinewave inputs, the cosine nonlinearity associated with phase-detecting ϕ_{ETF} using a multiplier driven by square waves represents a significant reduction in linearity. This is inherent to using a nonlinear multiplier as the ΣΔM's summing node.
- Two-step phase digitizers can first perform a coarse ϕ_{ETF} measurement using e.g. a single slope algorithm, and then run a fine conversion in which the phase references of the PDΣΔM are placed close to ϕ_{ETF}. This reduces quantization noise proportionally and nonlinearity cubically, thus greatly improving the performance of the phase digitizer.
- The remaining $T^{0.9}$ nonlinearity in ϕ_{ETF} can be removed in several ways, e.g. by duty-cycle modulation of the heater drive or by using a slightly non-linear decimation filter.
- New two-ETF or multi-ETF read-out architectures have been shown to enable self-referenced thermal diffusivity-based temperature sensing. In such architectures, both a temperature-dependent and temperature-independent signal are measured and processed to that their ratio is an accurate, linear function of absolute temperature.

The above represents several steps forward towards competitive ETF-based temperature sensors. However, future work will be required to actually achieve this goal.

6.2 Future research directions

6.2.1 Thermal management

6.2.1.1 ETFs in deep submicron CMOS technology

While this thesis has made progress in scaling ETFs, further work is required to implement ETFs in nanometer CMOS processes (e.g. 32, 28nm technologies and below), in which their performance should improve exponentially. Based on the current understanding of ETFs, scaling them down to such technologies is quite feasible, and their SNR should improve so as to outperform BJTs at technologies nodes of 45nm and below. Table 6.1 represents an estimate for the performance of a scaled ETF in 32nm, based on the modeling and measurement results presented in this thesis.

Table 6.1: Predicted ETF performance in deep submicron CMOS technology.

	Currently existing ETFs	Currently existing BJTs	32nm BJT	32nm ETF *(expected)*
Accuracy	0.2°C untrimmed	0.25°C batch trim	≈5°C untrimmed	**0.5°C untrimmed**
Power	2.5mW	**7.4µW**	1.6mW	500µW +circuit
Resolution	0.02°C_{rms}	**0.018°C**	0.45°C	0.2°C_{rms}
Conversion rate	0.16S/s	10S/S	1.2kS/s	**1kS/s**
Range	**-70°C to 225°C**	-70°C to 130°C	µP operation	µP operation
Area	4600µm^2 +circuit	0.12mm^2	0.02mm^2	**9µm^2** +circuit
Reference	This work	[6.1][6.2]	[6.3] [6.4]	

As this table shows, ETFs should have a conversion rate, resolution and power requirement similar to state-of-the art BJT-based temperature sensors, while offering about 10x lower inaccuracy. This should position them to be competitive to existing temperature sensors in those technologies.

While some thermal interference studies have been done in this work, the effect of switching on and off large blocks of digital circuitry in close proximity to the ETF will need to be studied in a more realistic implementation.

6.2.1.2 High-speed readout and associated readout circuitry

The performance of the ETF's readout circuit will have to keep up with the improvements expected from scaled ETFs in deep submicron technologies. This will most likely require new circuits, for example to achieve the much higher conversion rates typically required by thermal management (up to 10kS/S, now 1S/s). The PDΣΔM was selected for its ability to average offset and noise so that the ETFs could be accurately characterized, but the requirements on accuracy and resolution in thermal management applications are less critical than conversion rate and circuit area requirements. Therefore, the PDΣΔM may not be the best topology for ETF readout in thermal management; a phase digitizer based on successive approximation or a different delay-to-digital converter may be better suited for such high conversion rates.

For scaled ETFs, the phase error introduced by the finite bandwidth of the front-end of the phase digitizer will correspond to a proportionally more significant temperature error, especially at high f_{drive}. For topologies like the PDΣΔM described in chapter 4 or similar, this will drive up the current consumption of the transconductor, eventually overwhelming the power consumption of the now scaled ETF. Techniques such as direct demodulation [6.5] may be a way to break this tradeoff. Alternatively, a phase calibration or a phase autozeroing step is required before digitizing ϕ_{ETF}.

6.2.1.3 ETF improvements

Thermal management requires many silicon ETFs located across the die to be managed. Although their resolution should improve greatly by using ETFs with smaller geometries, their limited SNR will always be an important optimization target. Potential improvements in this direction could be achieved by optimizing the thermopile, e.g. to cover a larger fraction of the area around the ETF's heater, or by investigating other structures that may have a higher effective Seebeck coefficient. Another improvement could be to replace the thermopile with a more temperature-sensitive element, such as a BJT or an (n-well) resistor. Although such ETFs might have an increased untrimmed inaccuracy, the improvement in resolution and/or conversion rate might outweigh this drawback in some applications.

6.2.2 Self-referenced wide range temperature sensors

6.2.2.1 Alternative architectures

The self-referenced temperature sensor presented in chapter 3 and chapter 5 was based on measuring two ETFs. One ϕ_{ETF} was used as the temperature-dependent signal and the difference between two ϕ_{ETF}s , ϕ_{ox}, was used as the temperature-independent signal of a ratiometric temperature sensor. While measurement results show that the inaccuracy of this sensor, at ±0.4ºC (3σ), is very good after a single trim, there is a significant noise penalty associated with dividing ϕ_{ETF} by a relatively small ϕ_{ox} signal.

There are several other architectures that are worth exploring in the further development of self-referenced sensors. One would be to separate the measurement of ϕ_{Si} and ϕ_{ox} by using dedicated ETFs for each. The ϕ_{Si} ETF could be done using a relatively large ETF, having a good untrimmed inaccuracy. The two ETFs used to obtain ϕ_{ox} could be very small, since only their difference (=ϕ_{ox}) is critical for performance. By operating this readout at a higher f_{drive}, the absolute value of ϕ_{ox} can be increased, so that the ϕ_{Si}/ϕ_{ox} ratio can have a higher resolution. Moreover, if ϕ_{Si} is sufficiently

accurate, it can be used to trim for errors in ϕ_{ox} at room temperature, so that trimming using an external reference temperature sensor is no longer needed.

It may also be possible to measure ϕ_{ox} less frequently, since it is not very temperature dependent and the rate of on-chip transients is limited. This could increase the conversion rate and/or reduce power consumption.

A final self-referenced concept would move away from the paradigm in which ϕ_{ox} is established using two ETFs. Instead, an oxide-dominated ETF could be developed, i.e. an ETF of which ϕ_{ETF} is sufficiently temperature-independent to act as the time reference in the ratiometric sensor. Effectively, this means an ETF is required for which most ϕ_{ETF} contribution is from heat diffusing through oxide. This may yield improved D_{SiO2} measurement.

6.2.2.2 Improved D_{SiO2} measurement

In this work, the ETFs used to measure D_{SiO2} were based on extending an existing thermal path in silicon with an oxide component; both the field oxide / STI and the oxide part of deep isolation trenches have been explored. The trench oxide yielded high accuracy results, but requires SOI technology; the field oxide / STI ETFs were not explored in much details, since they were found to suffer from increased spread, having more than one degree of freedom.

There may be ways to measure D_{SiO2} more accurately or with higher resolution. For example, D_{SiO2} could be measured more directly by locating a relative temperature sensor directly underneath a polysilicon heater, or by implementing an ETF in the higher metal layers of a CMOS process, for example using thin-film resistors.

Alternatively, further characterization and modeling of field oxide ETFs may increase understanding of these ETFs so that they can be designed to achieve higher performance.

6.2.2.3 Fundamental operating range limitations

It may be worthwhile to further investigate the fundamental minimum and maximum operating temperature of ETF-based temperature sensors. Fundamentally, D_{Si} is strongly temperature dependent over a much wider range than has been explored so far, and e.g. operation at -100ºC or 300ºC may be of interest for space or industrial applications. At sufficiently high temperatures, other elements of the CMOS process, such as aluminum interconnect, may limit the operating range. It would be interesting to explore the potential of ETFs in other substrates / technologies, such as silicon carbide (SiC). Since the thermal diffusivity of all crystalline semiconductors is strongly temperature-dependent, several spin-off temperature sensors could be developed.

6.2.3 Other applications of ETFs and phase digitizers

6.2.3.1 Improved frequency references

As discussed in chapter 3, ETFs can be used to make accurate integrated frequency references. They output a constant frequency by locking a VCO to ϕ_{ETF}, and then compensate the temperature dependence of ϕ_{ETF} using a 'conventional' BJT-based temperature sensor. An accuracy of 0.1% over the over the -55ºC to 125ºC range has been achieved, and strong scaling properties have been shown [6.6]. However, the strong temperature dependence of D_{Si} fundamentally limits their performance. It might be possible to use D_{SiO2} instead, which should achieve an order of magnitude improvement.

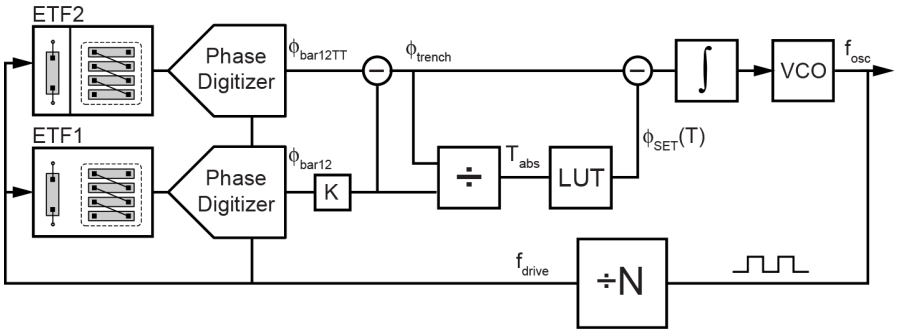


Fig. 6.1: Frequency reference based on D_{Si} and D_{SiO2}.

Another way is to use an architecture similar to the self-referenced ETF-based temp sensor. This is shown in Fig. 6.1. This loop locks to ϕ_{trench} that is made maximally flat over T using a gain factor K, and uses a ϕ_{ETF}/ϕ_{trench} absolute temperature sensor to compensate for the residual temperature dependence. Based on measurement results, a ±250 ppm frequency reference can be expected. This should improve further in a more optimized design. Such levels of inaccuracy will make thermal diffusivity-based frequency references competitive to crystal oscillators [6.6].

6.2.3.2 Measuring thermal conductivity using an ETF

Many integrated thermal sensors measure thermal conductivity and/or thermal diffusivity as a proxy for a physical variable of interest. Examples are flow sensors such as wind sensors [6.7], or gas sensors [6.8].

A thermal-conductivity-based gas sensor consists of an ETF deposited on a thin silicon layer bridging a cavity. In a well-designed sensor, the thermal conductivity of the gas inside the cavity strongly affects the thermal transfer impedance between the heater and the thermopiles, thus creating both variations in amplitude and phase shift characteristics. As for the ETFs discussed in this thesis, phase information tends to be more accurate, since the absolute gains and sensitivities of on-chip components suffer from process spread.

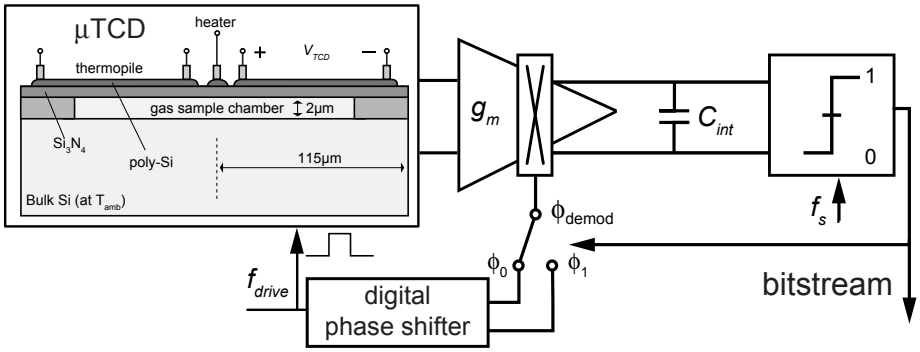


Fig. 6.2: PDΣΔM interfacing a thermal conductivity gas sensor.

A short study [6.9] attempts to interface the gas sensor with a PDΣΔM, to leverage the associated precision and resolution benefits. Fig. 6.2 shows the system. Simulations show that a resolution of $0.07\%_{rms}$ H_2 can be achieved for only 50µW power dissipation [6.9]. This would be state of the art, but further work is required to implement the above system and to measure its sensitivity to changing gas concentration.

6.2.3.3 Other applications

The use of phase-domain readout techniques can be extended to other (non-thermal) sensors that also output their information in the phase domain or can be (re-) designed to do so. This has already been applied to time-of-flight image sensors, i.e. 3D image sensors in which the time it takes for light to reflect back to a sensor is a measure for the distance between the camera and the object [6.10]. In a second example [6.11], a thermistor is read out by embedding it in a Wien-bridge filter, and then digitizing this filter's phase shift to measure temperature.

Another use for ETFs could be to improve the thermal characterization of a process technology. Using ETFs, the effective thermal conductivity and thermal diffusivity of integrated devices can be measured directly, which will enable more accurate design for self-heating and thermal gradients. This may be more precise and faster than current methods, which rely on long finite element simulations to achieve precision. An example can be found in the work of Altet [6.12], who uses ETFs to characterize thermal crosstalk in RF LNAs.

6.2.4 Summary

This thesis has discussed the development of integrated temperature sensors based on heat diffusion using electrothermal filters, and the precision readout circuits used to digitize their output. As evidenced by this chapter, several important steps forward have been taken, but the sections above still list a significant amount of future work. This future work will most likely be targeted towards temperature sensors for thermal management of microprocessors and integrated frequency references, together with a number of spin-off developments such as for the gas sensor outlined above.

6.3 References

[6.1] K. Souri, Y. Chae and and K.A.A. Makinwa, “A CMOS Temperature Sensor With a Voltage-Calibrated Inaccuracy of ±0.15°C (3σ) From -55 to 125°C,” *Digest ISSCC*, pp. 208 – 210, Feb. 2012.

[6.2] 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 to 130°C,” *Digest ISSCC*, pp. 342 – 343, Feb. 2009.

[6.3] Y.W. Li, H. Lakdawala, A.Raychowdhury, G. Taylor and K. Soumyanath, “A 1.05V 1.6mW, 0.45°C 3σ Resolution ΣΔ based Temperature Sensor with Parasitic Resistance Compensation in 32nm Digital CMOS Process, *IEEE Journal of Solid-State Circuits*, 2009.

[6.4] J. Shor, K. Luria, D. Zilberman, "Ratiometric BJT-based thermal sensor in 32nm and 22nm technologies", *ISSCC Digest,* pp. 210-212, Feb. 2012.

[6.5] S.H. Shalmany, “Thermal Diffusivity-Based Temperature Sensors for Microprocessors”, M.Sc. Thesis, Delft University of Technology, Delft, The Netherlands, Oct. 2010.

[6.6] S. M. Kashmiri, “Thermal-Diffusivity-Based Frequency References in Standard CMOS”, Ph.D. Thesis, Delft University of Technology, Apr. 2012.

[6.7] J. Wu, Y. Chae, C. van Vroonhoven and K.A.A. Makinwa, “A 50mW CMOS Wind Sensor with ±4% Speed and ±2° Direction Error,” *Digest ISSCC*, pp. 106 – 108, Feb. 2011.

[6.8] G. de Graaf, H. Wu and R. F. Wolffenbuttel, “Thermal Analysis, Fabrication and Signal Processing of Surface Micromachined Thermal Conductivity Based Gas Sensors”, *Proceedings IEEE EuroSimE 2011*, Linz, Austria, April 2011.

[6.9] C. van Vroonhoven, G. de Graaf and K.A.A. Makinwa, “Phase readout of thermal conductivity-based gas sensors,” *Proc. of the 4th International Workshop on Advances in Sensors and Interfaces (IWASI)*, pp. 199-202, June 2011.

[6.10] R.J. Walker, J.A. Richardson, R.K. Henderson, "A 128×96 pixel event-driven phase-domain ΔΣ-based fully digital 3D camera in 0.13μm CMOS imaging technology," *ISSCC Digest*, pp. 410-412, Feb. 2011.

[6.11] M. Shahmohammadi, K.Souri, K.A.A. Makinwa, "A resistor-based temperature sensor for MEMS frequency references," Proc. *ESSCIRC,* pp. 225-228, Sept. 2013.

[6.12] J. Altet et al., “Heterodyne lock-in thermal coupling measurements in integrated circuits: Applications to test and characterization”, Rev. Sci. Instrum, vol. 80, no. 2, 2009.

Summary

This thesis describes the theory, design and implementation of a new class of integrated temperature sensors, based on heat diffusion. In such sensors, temperature is sensed by measuring the time it takes for heat to diffuse through silicon. An on-chip thermal delay can be determined by geometry and the *thermal diffusivity* of silicon, and since the thermal diffusivity of crystalline silicon is strongly temperature-dependent, the thermal delay is also temperature dependent. The sensor structures that measure such delays are known as Electrothermal Filters (ETFs).

The field of temperature sensing using ETFs is still relatively unexplored. This thesis expands upon recent proof-of-concept research by studying the performance of ETF-based temperature sensors in more detail. Its main goal is to analyze whether or not ETFs can be used as competitive temperature sensors in CMOS technology.

The thermal delay in an ETF can be very well-defined, since the silicon used in IC fabrication is highly pure, and the lithographic inaccuracy with which planar structures can be made is constantly improving. ETFs thus scale along the trend of Moore's law, and they can become smaller, faster and more accurate in more advanced CMOS technologies.

The main application of ETF-based temperature sensors is the thermal management of microprocessors: modern microprocessors require tens of small, fast and accurate temperature sensors to prevent overheating and to dynamically allocate processing power. Chapter 1 outlines this and other applications in some detail, and shows why existing temperature sensors are not well-suited for these applications.

Chapter 2 presents the theoretical background to ETFs, starting with a discussion on the thermal diffusivity of silicon and silicon dioxide, analyzing their nominal values and their (different) temperature dependencies. It then discusses what type of ETFs to use to measure these most accurately. Furthermore, it outlines a family of performance metrics that enable a more formal study of ETFs characteristics. These are divided in two main categories: accuracy (e.g. error due to lithography, doping sensitivity) and resolution (how to optimize signal-to-noise ratio (SNR), advantages of scaling and using thermally isolated technologies such as silicon-on-insulator (SOI) processes).

Chapter 3 presents various systems that can be built around ETFs. Using either silicon ETFs, oxide ETFs or a combination of both, several systems can be built, each with different functionality. The two systems described in most detail are the single silicon-ETF temperature-to-digital converter, and the silicon and oxide ETF-based self-referenced temperature-to-digital converter. The former uses an accurate time reference (such as a crystal oscillator) to convert an ETF's thermal delay to an absolute temperature measurement. The latter performs a ratiometric temperature measurement and does not require an external time reference. This chapter also defines all of the circuit building blocks required to implement these systems, and derives their specifications from ETF theory and first principles. It concludes by showing that phase-domain sigma-delta modulators (PDΣΔMs) are the most appropriate readout circuits for precision ETF readout.

Chapter 4 discusses PDΣΔMs in more detail and presents their transistor-level implementation. PDΣΔMs are a class of time-to-digital converters that are well-suited to digitize the thermal delay contained in the small, noisy signal at the ETF's output. High precision and resolution are achieved by oversampling, noise shaping and the use of dynamic error correction techniques. Several methods of linearizing ETF-based temperature-to-digital converters are also presented. The chapter ends with

measurements on a standalone PDΣΔM, in order to show that its measurement error is sufficiently low to be able to characterize ETFs.

Chapter 5 presents the measurement results for the ETFs studied in this thesis. There are results on ETF inaccuracy and resolution as a function of geometry and process technology, as well as results on sensitivity to doping fluctuations, mechanical stress, thermal interference and self-heating. The measured differences in ETF performance in bulk CMOS and SOI CMOS technology are also presented and analyzed. The lowest measured untrimmed inaccuracy for a single-ETF sensor in 0.18μm CMOS technology was ±0.2ºC (3σ) from -55ºC to 125ºC. A self-referenced sensor, based on measuring the ratio of the thermal delay in silicon over that in oxide, shows a measured inaccuracy of ±0.4ºC (3σ) from -70ºC to 200ºC after a one-point trim.

Chapter 6 lists the main findings of this thesis and uses the measurement results to draw some conclusions on the competitiveness of ETFs. It shows that ETFs, due to their scalability, are attractive for thermal management of SoCs in deep submicron CMOS technologies. This thesis ends with some recommendations for further research.

Samenvatting

Dit proefschrift beschrijft de theorie, het ontwerp en de implementatie van een nieuw type geïntegreerde temperatuursensor, dat is gebaseerd op warmtediffusie. Dit soort sensors meet temperatuur door de snelheid te meten waarmee warmte door silicium diffundeert. Een "thermische vertraging" op een chip wordt bepaald door geometrie en door de *thermische diffusiviteit* van silicium. Omdat deze erg temperatuurafhankelijk is, kan een dergelijke thermische vertraging gebruikt worden als basis voor een temperatuursensors. De sensorstructuren die de thermische vertraging meten staan bekend als elektrothermische filters (ETFs).

Het gebruiken van ETFs om temperatuur te meten is als onderzoeksgebied nog relatief onverkend. Na een aantal recente 'proof-of-concepts' worden ETF-gebaseerde temperatuursensoren in dit proefschrift in meer detail bestudeerd. Het hoofddoel van dit proefschrift is om te analyseren of ETFs als temperatuursensor kunnen concurreren met de reeds bestaande sensoren in CMOS technologie.

De thermische vertraging in een ETF kan erg nauwkeurig worden gedefinieerd, omdat het silicium dat wordt gebruikt bij de fabricage van IC's zeer zuiver is en omdat planaire structuren door de steeds beter wordende lithografie steeds nauwkeuriger kunnen worden gefabriceerd. Op deze manier volgen ETF-gebaseerde temperatuursensoren een soort wet van Moore, doordat ze kleiner, sneller en nauwkeuriger worden in meer geavanceerde CMOS procestechnologieën.

De belangrijkste toepassing van ETF-gebaseerde temperatuursensoren is het thermisch beheer van microprocessoren. Moderne microprocessoren worden door hun hoge energiegebruik nogal warm, en elke chip heeft tientallen kleine, snelle, maar ook nauwkeurige temperatuursensoren nodig om oververhitting te voorkomen en om het energieverbruik over meerdere processorcores te verdelen. In hoofdstuk 1 worden deze en andere toepassingen beschreven, en wordt ook duidelijk gemaakt waarom bestaande temperatuursensoren hiervoor minder geschikt zijn.

Hoofdstuk 2 vormt de theoretische achtergrond voor ETFs en begint met een korte uiteenzetting over de thermische diffusiviteit van silicium en siliciumdioxide, waarbij de nominale waarden en de temperatuurafhankelijkheid worden geanalyseerd. Vervolgens wordt beschreven met welke soort ETFs de thermische diffusiviteit het meest nauwkeurig kan worden gemeten. Daarnaast wordt er een opsomming van prestatie-indicatoren gemaakt, aan de hand waarvan de eigenschappen van ETFs vervolgens op een meer gestructureerde wijze kunnen worden bestudeerd. Er zijn hierbij twee hoofdcategoriën: nauwkeurigheid (bijv. gevoeligheid voor lithografische fouten of doteringsvariatie) en resolutie (bijv. hoe de signaal-ruisverhouding kan worden geoptimaliseerd en de voordelen van het verkleinen van ETFs en/of het gebruik maken van een thermisch geïsoleerd substraat).

In hoofdstuk 3 wordt een aantal ETF-gebaseerde systemen besproken. Er zijn verschillende functies te realiseren door systemen die gebruik maken van oftewel ETFs in silicium, ETFs in silicumdioxide of een combinatie van beide. De twee systemen die het meest uitvoerig worden besproken zijn de temperatuur-naar-digitaalomzetter die is gebaseerd op één silicium ETF en de sensor waarin de verhouding tussen de thermische diffusiviteit van silicium en siliciumdioxide wordt bepaalt. Het eerste systeem maakt gebruik van een externe tijdsreferentie (zoals bijvoorbeeld een kristaloscillator) om de faseverschuiving van de ETF te koppelen aan absolute temperatuur. Het tweede systeem verricht een ratiometrische tijdsmeting en heeft daarom geen externe tijdsreferentie nodig. Verder wordt er in dit hoofdstuk een inventarisatie gemaakt van welke schakelingsbouwstenen er nodig zijn om deze systemen te implementeren en worden de specificaties van deze bouwstenen opgesteld. Het hoofdstuk concludeert dat fase-domein sigma-delta modulatoren (PDΣΔMs) het meest geschikt zijn om ETFs met hoge nauwkeurigheid uit te lezen.

Hoofdstuk 4 beschrijft PDΣΔMs in meer detail en laat zien hoe deze geïmplementeerd kunnen worden op transistorniveau. PDΣΔMs zijn een type tijd-naar-digitaalomzetters dat zeer geschikt is voor het digitaliseren van de thermische vertraging die het kleine, ruizige signaal aan de uitgang van de ETF heeft. Door oversampling, ruisverplaatsing en door gebruik te maken van dynamische foutcorrectiemethoden wordt zowel een hoge resolutie als een hoge nauwkeurigheid behaald. Verder wordt een aantal methodes om ETF-gebaseerde temperatuur-naar-digitaalomzetters te linearizeren besproken. Het hoofdstuk besluit met meetresultaten die laten zien dat een PDΣΔM zónder ETF voldoende nauwkeurig is om gebruikt te kunnen worden bij het karakteriseren van ETFs.

Hoofdstuk 5 presenteert de meetresultaten behorende bij de ETFs uit dit proefschrift. Er zijn resultaten ten aanzien van de onnauwkeurigheid en de resolutie van ETFs als functie van hun geometrie en het proces waarin ze worden gefabriceerd, alsook ten aanzien van de invloed van variatie in dotering, mechanische stress, thermische interferentie en zelfopwarming. De gemeten verschillen tussen ETFs in bulk CMOS en SOI CMOS worden ook gepresenteerd. Voor de sensor met één ETF en een tijdsreferentie is de kleinst gemeten ongetrimde onnauwkeurigheid gelijk aan ±0.2ºC (3σ) over een bereik van -55ºC tot 125ºC, voor ETFs in een 0.18µm bulk CMOS proces. Voor de sensor die de verhouding tussen the thermische vertraging in silicium en siliciumdioxide meet is de kleinste onnauwkeurigheid gelijk aan ±0.4ºC (3σ) van -70ºC tot 200ºC, na een 1-puntstrim.

In hoofdstuk 6 worden de belangrijkste resultaten uit dit proefschrift opgesomd en wordt er op basis van de meetresultaten een conclusie getrokken over het concurrerend vermogen van ETFs. Het hoofdstuk laat zien dat ETFs door hun schaalbaarheid een aantrekkelijke optie zijn voor het thermisch managen van SoCs in nanometer CMOS technologiën. Het proefschrift rondt af met een aantal aanbevelingen ten aanzien van toekomstig onderzoek.

List of Publications

Journal papers:

C.P.L. van Vroonhoven and K.A.A. Makinwa, “An SOI thermal-diffusivity-based temperature sensor with ±0.6°C (3σ) untrimmed inaccuracy from −70°C to 225°C”, *Sensors and Actuators A: Physical*, Volume 188, pp. 66-74, Dec 2012.

Conference papers:

C. van Vroonhoven, D. D'Aquino, K. Makinwa, "A ±0.4°C (3σ) −70 to 200°C time-domain temperature sensor based on heat diffusion in Si and SiO_2," *IEEE ISSCC Digest,* pp. 204-206, Feb 2012.

Jianfeng Wu, C. van Vroonhoven, Youngcheol Chae, K. Makinwa, "A 25mW CMOS sensor for wind and temperature measurement," *proc. IEEE Sensors,* pp. 1261-1264, Oct. 2011.

C.P.L. van Vroonhoven, K. A. A. Makinwa, "Thermal diffusivity sensing: A new temperature sensing paradigm," *proc. IEEE Custom Integrated Circuits Conference (CICC),* pp. 1-6, Sept. 2011.

C. van Vroonhoven, G. de Graaf, K. Makinwa, "Phase readout of thermal conductivity-based gas sensors," *IEEE International Workshop on Advances in Sensors and Interfaces (IWASI),* pp. 199-202, June 2011.

C. P. L. van Vroonhoven, K. A. A Makinwa, "An SOI thermal-diffusivity-based temperature sensor with ±0.6°C (3σ) untrimmed inaccuracy from −70°C to 170°C," *proc. International Solid-State Sensors, Actuators and Microsystems Conference (TRANSDUCERS),* pp. 2887 - 2890, June 2011.

Jianfeng Wu, Youngcheol Chae, C.P.L. van Vroonhoven, K.A.A Makinwa, "A 50mW CMOS wind sensor with ±4% speed and ±2° direction error," *IEEE ISSCC Digest,* pp. 106-108, Feb. 2011.

C.P.L. van Vroonhoven, D. D'Aquino, 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," *IEEE ISSCC Digest,* pp. 314-315, Feb. 2010.

C. van Vroonhoven, M. Kashmiri, K. Makinwa, "CMOS temperature sensors based on thermal diffusion," *International Workshop on Thermal Investigations of ICs and Systems (THERMINIC),* pp. 140-143, Oct. 2009.

C. P. L. van Vroonhoven, K. A. A. Makinwa, "Linearization of a thermal-diffusivity-based temperature sensor," *proc. IEEE Sensors,* pp. 1697-1700, Oct. 2009.

C. van Vroonhoven, K. Makinwa, "Thermal diffusivity sensors for wide-range temperature sensing," *proc. IEEE Sensors,* pp. 764 - 767, Oct. 2008.

C. P. L. van Vroonhoven, K. A. A. Makinwa, "A CMOS Temperature-to-Digital Converter with an Inaccuracy of ±0.5°C (3σ) from -55 to 125°C," *IEEE ISSCC Digest,* pp. 576 - 637, Feb. 2008.

Patents:

US8013636 Synchronous Phase Detection Circuit
US8896474 Two-stage phase digitizer
US8870454 Multiple electrothermal-filter device

Awards and Recognition:

Transducers 2011 Outstanding Paper Award
ISSCC 2008 Jan van Vessem Award for Outstanding European Paper
IEEE Sensors 2008 Best Student Paper Award (2nd place)
Simon Stevin Leerling 2008

About the Author

Caspar van Vroonhoven was born in The Hague, The Netherlands in April 1983. Soon after, he took apart the baby intercom, was gifted his first soldering iron and became an amateur radio operator. His original passion suppressed by puberty and adolescence, it took until May 2007 before he received his M.Sc. / ir. Degree from Delft University of Technology. His graduation project at the Electronic Instrumentation Laboratory was on integrated temperature sensors based on Electrothermal Filters. Supported by National Semiconductor Cooperation (later Texas Instruments), he continued this research towards his Ph.D. degree at the same laboratory.

His research has led to several patent applications and been recognized through several awards, among them the Transducers 2011 Outstanding Paper Award and the ISSCC 2008 Jan van Vessem Award for Outstanding European Paper.

After finishing his laboratory work at University in December 2011, he joined Texas Instruments in Freising, Germany in February 2012, where he worked on isolated current sensing as an analog design engineer. Since November 2013, he is working as a design engineer for Linear Technology in Ismaning, Germany.

Acknowledgements

There are many people without whom this thesis would not have come into existence. This page is but an attempt to express my gratitude to every one of them.

My first and foremost thanks go to Kofi Makinwa, my promoter. His unfaltering guidance would inspire during the ups and encourage during the inevitable downs of a research project that would span many years. We share a creativity and a curiousity that would drive our results, combining carefully planned 'next steps' with increasingly daring academic embodiments of 'shoot first, ask questions later'. Kofi's dedication to both the personal success of his team and the technical success of their results make him an excellent professor. I greatly value and will forever reap the benefits of having been a member of this team.

Dan D'Aquino at National Semiconductor (now Texas Instruments) had the challenging task of supporting a university project from a business line's perspective; a task he managed very well. I appreciate both the technical and the professional insights that he shared, as well as his effort to steer the project to commercial relevance without limiting academic freedom. He also taught me how to surf (or at least tried to), which is yet another reason why it's a pity that we didn't work together more at TI than we did. Thanks also to Bijoy Chatterjee for supporting this project and supporting this student as the always friendly, supportive and well-connected university liason for TI that he is. Thanks to Wilko Kindt and colleagues at TI's Delft Design Centre for technical and logistical support. Phyllis Lee, JR Servando Aguilar and Wen-Jung Hsu did most of the IC layout work for the SOI testchip, for which I am grateful.

Thanks go to my fellow (former) Ph.D. students: Mahdi, Kamran, Rong, Qinwen, Ugur, André, Zhichao, Saleh, Muhammed, Fabio, Jiawei, Sha, Zili, Ning and Junfeng; for accepting this Dutch guy in their midst, as well as being part of an inspiring research lab. Mahdi and I were able to learn from each other and collaborate on our similar projects, and I'm glad this was to be the basis for our friendship. It was interesting to work together with several master students, including Sijia, Shagun, Jianfeng, Ge, Gaurav, Navid, Shishir and Jeroen. Thanks also go to Youngcheol and Michiel, for being excellent sparring partners and sources of new ideas.

The Electronic Instrumentation Laboratory has been a great place to study and work. Thanks go to all the staff members and all the people that haven't already been mentioned above. Thanks to Piet, Ger, Jeroen, Maureen en Zu-Yao. I'd like to thank Joyce Siemers in particular for her support.

There have been many others that have supported, playfully doubted, but ultimately encouraged this work. My friends and my extended family have always shown great interest in to what is, in fact, a pretty specialized and highly technical topic. I can only hope that I've been and will be able to return in kind.

This thesis is dedicated to those who have perhaps cared the most about its completion. In their absence I feel even more grateful and blessed for having such a supportive family and such a very supportive girlfriend, Tammy. Their love and understanding are at the foundation of all I do.

Caspar van Vroonhoven


TUDelft