Low-cost temperature sensors in CMOS technology

Yousefzadeh, B.

DOI
10.4233/uuid:389f6a1c-f86b-48bd-a53c-f66bea5a3db8

Publication date
2024

Document Version
Final published version

Citation (APA)
Yousefzadeh, B. (2024). *Low-cost temperature sensors in CMOS technology*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:389f6a1c-f86b-48bd-a53c-f66bea5a3db8

LOW-COST TEMPERATURE SENSORS IN CMOS TECHNOLOGY

LOW-COST TEMPERATURE SENSORS IN CMOS TECHNOLOGY

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus Prof. dr. ir. T. H. J. J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on
Thursday 30 May 2024 at 15:00

by

Bahman YOUSEFZADEH

Master of Science in Electrical Engineering, Sharif University of Technology, Iran
born in Shoshtar, Iran

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

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

Independent members:

Prof. dr. ir. F. P. Widdershoven	Delft University of Technology
Prof. dr. Y. Chae	Yonsei University, South Korea
Dr. ir. M. A. P. Pertijs	Delft University of Technology
Dr. F. Sebastiano	Delft University of Technology
Dr. J. Shor	Bar-Ilan University, Israël
Prof. dr. ir. A. J. P. Theuwissen	Delft University of Technology

Keywords: CMOS, BJT, Accurate, Low-cost, Calibration, Heater-assisted, On-chip heater, Multi-function sensor, ADC

Printed by: ProefschriftMaken

ISBN 978-94-6384-590-8

An electronic version of this dissertation is available at
http://repository.tudelft.nl/

To Joëlle
and my parents

Contents

Chapter 1

Introduction

Temperature greatly affects the performance of integrated circuits (ICs), and so temperature sensors are often necessary to ensure reliable operation and stable performance. For instance, they can be used to detect hot-spots in large system-on-chips (SoCs) to prevent them from premature failure and increase their performance. They are also used to compensate for the temperature dependence of other sensors and references and thus increase their accuracy. Depending on the application, the required sensing range varies: from small (e.g., ± 10 °C around body temperature) in biomedical sensors [1.1], to large (e.g., -40 °C to $+150$ °C) in automotive electronics [1.2]. The required inaccuracy also varies with application: from ± 1 °C for the thermal management of SoCs [1.3], to less than ± 0.1 °C in precision sensors [1.4],[1.5], and references [1.6].

Temperature can be determined by measuring an electrical parameter of an on-chip element with a known temperature dependency. To make an accurate temperature sensor, the variation in this parameter due to the fabrication process and the mechanical stress of packaging must be corrected by calibration or by design. In principle, parameter variation can be sufficiently suppressed by using multiple calibration points. However, this will proportionally increase production costs.

1.1 Background

The temperature behavior of Bipolar Junction Transistors (BJTs) has been well studied [1.7]–[1.9], and shown to be stable over long periods of time [1.10]. Furthermore, the effect of process spread in BJTs can be corrected by a single calibration. As a

result, BJTs are often used in integrated temperature sensors.

In early bipolar technologies, the difference between the Base-Emitter voltages of two BJTs with different current densities (ΔV_{BE}) was used to generate stable bias currents [1.11]. Later, temperature sensors based on this principle were introduced. These typically had outputs in the form of voltages [1.12], currents [1.13],[1.14], or frequencies [1.15]. Analog-to-digital converters (ADCs) were subsequently added to realize temperature-to-digital converters (TDCs) [1.16], whose digital outputs were more robust and could be easily connected to microprocessors.

TDCs then evolved by shifting from bipolar technologies to CMOS, mainly to be compatible with the mainstream digital circuits. Fortunately, BJTs can still be realized in CMOS technologies as lateral [1.17], or vertical [1.18] devices. MOSFETs also exhibit bipolar-like behavior when biased in the weak inversion region [1.19].

1.2 Production cost of a temperature sensor

The cost of producing a temperature sensor can be divided into three parts: design, fabrication, and testing. Design accounts for the smallest fraction of the total costs, as it is a fixed amount that will be spread over a large number of samples. Fabrication costs are higher, as they consist of a one-time cost for the production of masks and then the repeated cost of fabricating each batch of wafers, which contain thousands of samples. The third and highest costs are due to testing, which may need to be carried out on each sample. Delivering a faulty sample can cause a larger system to fail, leading to even higher costs.

Temperature sensors can be tested at wafer-level, package-level, or both. During wafer-level testing, non-functional samples are marked and then excluded from the packaging. A performance assessment is also carried out, which is then used to sort samples into different bins with different performance ranges. When required, calibration and trimming are carried out during wafer-level testing. During package-level testing, errors that might arise from either erroneous wire-bonding, or the high pressure used during the encapsulation process, will also be detected.

The cost of wafer-level testing is usually lower than that of package-level testing, because thousands of samples can be tested at the same time. A test probe then scans the wafer, makes connections to each sample, and performs an electrical test. For temperature calibration, the wafer is placed on a temperature-controlled chuck. In a minute-long process, the wafer and the chuck reach a thermal equilibrium after which the electrical measurements can start. Due to its higher costs and infrastructure

requirements, package-level testing is usually limited to simple functionality tests.

1.3 Motivation and objectives

At the start of this work, a state-of-the-art temperature sensor in the target CMOS technology (0.16-μm) could obtain 3σ-inaccuracies of ± 0.15 °C from -55 °C to 125 °C [1.20]. However, this shifted somewhat over two production batches, and could only be corrected at the expense of more calibration. In earlier 0.7-μm CMOS technology, by contrast, this shift was negligible [1.10]. Table 1.1 summarizes the results of [1.20] for two different production batches, and in both ceramic and plastic packaging. To correct the shift between the two batches, a batch-specific set of calibration parameters (A, B, α) had to be determined. The same procedure had to be repeated when the sensor was packaged in plastic. The need to re-calibrate separate batches and differently packaged samples made it impractical to use the sensor in [1.20] as a low-cost precision product.

The goal of the work described in this thesis is the realization of a temperature sensor that does not require batch-calibration. To this end, the first challenge is to determine the root cause behind the batch-to-batch shift observed in [1.20]. From the reported performance of this sensor (see Table 1.1), it could not be conclusively determined whether this shift was due to the available BJTs and resistors, or due to an architectural limitation. The second challenge is to develop low-cost methods to correct the sensor's shift in different packages.

1.4 Method

To address the first challenge, an accurate and process-insensitive sensor was designed, which unlike [1.20] does not require individual batch-calibration, and thus, can be produced with lower costs. This was achieved by adding circuit techniques to the sensor's front-end to mitigate its sources of error. The accuracy of the sensor was then verified by measuring samples from three different batches. This approach enabled the experimental characterization of the main sources of spread, thus capturing physical effects not included in the device models.

To address the second challenge, low-cost calibration methods were developed to help reduce test costs. These calibration methods involved measurements at two different temperatures. Room temperature was used as one temperature point, while

	Batch-1	Batch-2	Batch-2
CMOS technology (μm)	0.16	0.16	0.16
Packaging type	Ceramic	Ceramic	Plastic
Temperature range (°C)	−55 to +125	−55 to +125	−55 to +125
Untrimmed inaccuracy (3σ)	±0.6 °C	±0.6 °C	±0.8 °C
Trimmed inaccuracy (3σ)	±0.15 °C	±0.25 °C	±0.25 °C
α	15.44	15.45	15.47
A	613.31	610.74	611.59
B	283.70	282.93	283.94

Table 1.1: Performance summary of the state-of-the-art CMOS BJT-based temprature sensor [1.21]. α, A, and B are average calibration parameters.

an on-chip heater was used to elevate die temperature and create a second temperature point. The effectiveness of these heater-assisted calibration methods was then evaluated on sensors from different production batches and in different packages.

In the final step, additional functionality was added to the sensor by re-configuring its front-end circuitry, and so, making more cost-effective use of its area. The result is a multi-function sensor that can digitize external voltages and capacitances as well as temperature.

1.5 Organization of this thesis

Chapter 2 discusses four different methods for realizing a TDC in CMOS technology based on resistors, MOSFETs, thermal-diffusivity (TD), and BJTs. For each method, the sensing principle is explained, and then examples of typical readout circuits and the resulting inaccuracy are provided. This gives an overview of the expected inaccuracy of different TDCs and their calibration requirements.

Chapter 3 focuses on BJT devices. It starts with a discussion of the equations that govern the operation of ideal BJTs. Non-ideal effects that cause temperature inaccuracy, such as finite current gain and non-zero series resistances, are explained in detail. Lastly, the effect of mechanical stress on the BJTs is discussed.

Chapter 4 discusses the first TDC implementation in this thesis, which was intended to reduce the batch-to-batch shift observed in [1.20]. It starts with an analysis of the error sources in the sensor's front-end and then describes ways of suppressing them. Compared to the state-of-the-art, the sensor achieved 2× better accuracy, which also eliminated the need for batch calibration. Voltage calibration as a low-

cost alternative to temperature calibration was also verified for this sensor.

Chapter 5 describes the second TDC implementation in this thesis. It includes the realization of two TDCs that were equipped with on-chip heaters. These TDCs were used to investigate two different heater-assisted calibration methods. The effect of the mechanical stress of plastic packaging on the TDC, and low-cost methods to correct for this are also discussed in this chapter.

Chapter 6 describes the third TDC implementation in this thesis, which expanded the sensor's functionality. It describes how the circuitry of the TDC in Chapter 4 can be re-used to also digitize external voltages and capacitances.

Chapter 7 summarizes the main findings in this thesis and provides some suggestions for future work.

Bibliography

[1.1] A. Vaz, A. Ubarretxena, I. Zalbide, D. Pardo, H. Solar, A. Garcia-Alonso, and R. Berenguer, "Full passive UHF tag with a temperature sensor suitable for human body temperature monitoring," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 57, no. 2, pp. 95–99, 2010.

[1.2] R. W. Johnson, J. L. Evans, P. Jacobsen, J. R. Thompson, and M. Christopher, "The changing automotive environment: high-temperature electronics," *IEEE Transactions on Electronics Packaging Manufacturing*, vol. 27, no. 3, pp. 164–176, 2004.

[1.3] U. Sönmez, F. Sebastiano, and K. A. A. Makinwa, "Compact thermal-diffusivity-based temperature sensors in 40-nm CMOS for SoC thermal monitoring," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 3, pp. 834–843, 2017.

[1.4] M. H. Perrott, J. C. Salvia, F. S. Lee, A. Partridge, S. Mukherjee, C. Arft, J. Kim, N. Arumugam, P. Gupta, S. Tabatabaei, S. Pamarti, H. Lee, and F. Assaderaghi, "A temperature-to-digital converter for a MEMS-based programmable oscillator with $< \pm 0.5$-ppm frequency stability and $<$ 1-ps integrated jitter," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 1, pp. 276–291, 2013.

[1.5] S. Heidary Shalmany, D. Draxelmayr, and K. A. A. Makinwa, "A ± 36-μA integrated current-sensing system with a 0.3% gain error and a 400-μA offset from -55 °C to $+85$ °C," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 4, pp. 1034–1043, 2017.

[1.6] S. M. Kashmiri, M. A. P. Pertijs, and K. A. A. Makinwa, "A thermal-diffusivity-based frequency reference in standard CMOS with an absolute inaccuracy of $\pm 0.1\%$ from -55 °C to 125 °C," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 12, pp. 2510–2520, 2010.

[1.7] Y. P. Tsividis, "Accurate analysis of temperature effects in $I_C - V_{BE}$ characteristics with application to bandgap reference sources," *IEEE Journal of Solid-State Circuits*, vol. 15, no. 6, pp. 1076–1084, 1980.

[1.8] G. C. M. Meijer and K. Vingerling, "Measurement of the temperature dependence of the $I_C - V_{BE}$ characteristics of integrated bipolar transistors," *IEEE Journal of Solid-State Circuits*, vol. 15, no. 2, pp. 237–240, 1980.

[1.9] G. Wang and G. C. Meijer, "The temperature characteristics of bipolar transistors fabricated in CMOS technology," *Sensors and Actuators A: Physical*, vol. 87, no. 1, pp. 81 – 89, 2000.

[1.10] G. Wang, A. Heidari, K. A. A. Makinwa, and G. C. M. Meijer, "An accurate BJT-based CMOS temperature sensor with duty-cycle-modulated output," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 2, pp. 1572–1580, Feb. 2017.

[1.11] R. Widlar, "Some circuit design techniques for linear integrated circuits," *IEEE Transactions on Circuit Theory*, vol. 12, no. 4, pp. 586–590, 1965.

[1.12] R. Dobkin, "Monolithic temperature transducer," in *IEEE International Solid-State Circuits Conference (ISSCC). Dig. of Tech. Papers*, vol. XVII, 1974, pp. 126–127.

[1.13] M. P. Timko, "A two-terminal IC temperature transducer," *IEEE Journal of Solid-State Circuits*, vol. 11, no. 6, pp. 784–788, 1976.

[1.14] G. C. M. Meijer, "An IC temperature transducer with an intrinsic reference," *IEEE Journal of Solid-State Circuits*, vol. 15, no. 3, pp. 370–373, 1980.

[1.15] D. Van Maaren, J. Klijn, and G. C. M. Meijer, "An integrated micropower low-voltage temperature-controlled oscillator," *IEEE Journal of Solid-State Circuits*, vol. 17, no. 6, pp. 1197–1201, 1982.

[1.16] G. C. M. Meijer, A. J. M. Boomkamp, and R. J. Duguesnoy, "An accurate biomedical temperature transducer with on-chip microcomputer interfacing," *IEEE Journal of Solid-State Circuits*, vol. 23, no. 6, pp. 1405–1410, 1988.

[1.17] P. Krummenacher and H. Oguey, "Smart temperature sensor in CMOS technology," *Sensors and Actuators A: Physical*, vol. 22, no. 1, pp. 636 – 638, 1990.

[1.18] A. Bakker and J. H. Huijsing, "Micropower CMOS temperature sensor with digital output," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 7, pp. 933–937, Jul 1996.

[1.19] E. A. Vittoz and O. Neyroud, "A low-voltage CMOS bandgap reference," *IEEE Journal of Solid-State Circuits*, vol. 14, no. 3, pp. 573–579, 1979.

[1.20] K. Souri, Y. Chae, and K. A. A. Makinwa, "A CMOS temperature sensor with a voltage-calibrated inaccuracy of $\pm 0.15^{\circ}$C (3σ) from -55°C to 125°C," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 1, pp. 292–301, Jan. 2013.

Chapter 2

CMOS temperature-to-digital converters

This chapter gives an overview of various types of CMOS temperature-to-digital converters (TDCs). For each TDC type, its operating principle and some typical readout circuitry will be discussed. Based on the available literature, the achievable temperature inaccuracy, calibration requirements, and limitations will also be discussed, which then allows the TDCs to be compared. This chapter consists of five sections. The first four sections discuss TDCs based on resistors, thermal diffusivity (TD), MOSFETs, and BJTs, respectively. Section 5 concludes this chapter.

2.1 Resistor-based TDCs

2.1.1 Resistors in CMOS technology

When the behavior of a resistor with respect to temperature is known, its value can be digitized and used as a measure of temperature. The available resistors in standard CMOS technology are shown in Table 2.1, tohether with their associated parameters such as Sheet Resistance R_{sh}, Temperature Coefficient (TC), Voltage Coefficient (VC), and also a qualitative description of their flicker noise.

Metal resistors are made with the same material used to realize the interconnects in a chip. Initially, Aluminium (Al) or an alloy of Al with Silicon (Si) or Copper (Cu) was used as the interconnecting material in CMOS technology. Later, Cu was chosen for its lower resistivity compared to Al, its higher resistance to electro-migration, and also its better reliability. Metal resistors have a very low sheet resistance which, depending on the thickness of the metal layer used, ranges from ~ 0.02 to $\sim 0.08\ \Omega/\square$.

Resistor type	R_{sh} [a](Ω/□)	TC [b](%/°C)	VC [c](ppm/V)	$1/f$ noise
Metal	0.02 to 0.08	+0.3 to +0.4	Negligible	Negligible
N-well	1 k to 10 k	+0.3	∼10 k	Very low
Diffusion ($p+/n+$)	50 to 200	+0.08 to +0.15	100 to 500	Very low
Silicided diffusion	5 to 10	+0.25 to +0.4	Lower than diffusion	Low
Polysilicon ($p+/n+$)	50 to 350 [d]	−0.15 to +0.01	∼100	High
Silicided polysilicon	5 to 10	+0.25 to +0.4	Lower than polysilicon	Low

[a] Sheet resistance [b] Temperature coefficient [c] Voltage coefficient
[d] R_{sh} of high-resistive poly (undoped) is $\sim 1\ k\Omega/\square$

Table 2.1: Overview of resistor characteristics in CMOS technologies.

They have a large TC of $\sim 0.3\%/°C$ and little to no flicker noise. Metal resistors also show a very low dependence on applied voltages. This is mainly because, unlike diffusion or n-well resistors, their geometry hardly changes in response to applied voltages. Due to their large TC, decent linearity, and negligible low-frequency noise, metal resistors are very suitable candidates for realizing a TDC. However, due to their relatively low sheet resistance, these advantages come at the expense of area.

In standard single-well CMOS technology, *N-well resistors* are made with low-doped (n-type) silicon in a p-type substrate. In twin-well technologies, it is also possible to realize p-well resistors. Well resistors typically have a high R_{sh}, which can range from ~ 1 to ~ 10 k$\Omega/\square$. They exhibit a large TC of $\sim +0.3\%/°C$, and also very low flicker noise [2.1]. However, well resistors are known for their poor non-linearity or equivalently large voltage dependence. The resistance of an n-well resistor is modulated by the potential difference between the well and the underlying p-substrate. This is because the width of the depletion region formed in the reverse-biased junction between the p-substrate and the n-well varies with the voltage across the resistor. As a result, the effective geometry of the well-resistor varies and cause resistance variation [2.2]. The resistance of an n-well increases with an increase in the applied voltage, since a higher voltage extends the depletion region into the resistor region and reduces its effective thickness. Due to the relatively low doping concentration of the n-well, the depletion region predominantly extends inside the n-well rather than the p-substrate, which accounts for their relatively higher VC compared to diffusion resistors. Unless it is necessary to realize very large resistors in a small area, this makes them unsuitable for use in most analog circuits. Besides their high VC, they are also vulnerable to substrate noise, which can couple to the resistor via the well to substrate junction capacitance, and they are very sensitive to mechanical stress.

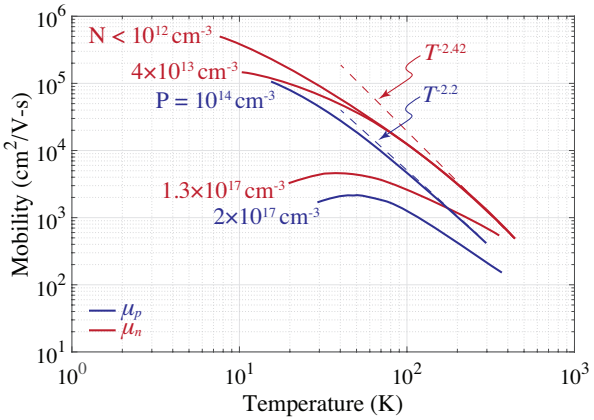


Figure 2.1: Temperature dependence of mobility in doped silicon for different doping concentrations.

Diffusion resistors are made from the n+ or p+ regions which form the source and drain of MOSFETs. These regions are usually silicided to provide a robust connection to the metal interconnect. Using unsilicided diffusion, a decent R_{sh} can be obtained, which depending on the doping concentration can range from $\sim$ 50 to $\sim$ 200 $\Omega/\square$. The TC of diffusion resistors depends on the doping concentration and ranges from $\sim$ +0.08%/°C to $\sim$ +0.15%/°C. As shown in Fig. 2.1, a higher doping level results in a lower TC since the dependence of mobility on temperature becomes weaker at higher doping levels. The TC of diffusion resistors is positive at all doping levels since the charge carrier mobility is mainly determined by phonon scattering [2.3]. Similar to n-well resistors, diffusion resistors have low-levels of flicker noise and suffer from a relatively large VC. The VC of diffusion resistors, however, is lower than that of well resistors [2.4]. This is due to their relatively higher doping concentration, which results in a thinner depletion region.

Poly resistors are made from the polysilicon material used to realize the gate of MOSFETs. Polysilicon consists of amorphous silicon, i.e., grains of silicon with different crystal orientations. As a result, the behavior of polysilicon resistors differs from that of resistors made from mono-crystalline silicon such as n-well or diffusion resistors. They exhibit the largest flicker noise compared to other resistors [2.5]. Poly resistors depend less on the applied voltage and are very linear since they are placed on top of an insulating layer of oxide. This prevents direct junctions between the resistor and the substrate. However, the bottom part of this resistor can still be depleted by the field across the oxide [2.6], which accounts for a small VC. Typical values for the VC of polysilicon resistors are $\sim$ 100 ppm/V. It is known that the average size

of the grains in the polysilicon plays a key role in determining their R_{sh} and TC [2.7]. As a result, the R_{sh} and TC of poly resistors are functions of the parameters that determine the size of these grains, such as the chosen time and temperature used to grow and, subsequently, anneal the polysilicon film [2.7]. Polysilicon resistors are often doped with p or n impurities, which can then cause their R_{sh} and TC to vary depending on the type and dose of the impurity used. It is also known that the dimension of the doped polysilicon resistors is a determining factor of their TC [2.3]. Being dependent on several parameters, the R_{sh} of poly resistors can range between $\sim$ 50 to $\sim$ 350 $\Omega/\square$, while their TC can vary from negative values, e.g., ~ -0.15 %/°C to positive values, e.g., $\sim +0.01$ %/°C [2.7]. As a result, they can be used as temperature-sensing elements in TDCs, in which a high TC is desirable, and also in precision analog circuits, where a near-zero TC is preferred.

There are two other types of resistors available in standard CMOS which are essentially variants of the ones already described: *silicided poly* and *silicided diffusion resistors*. Silicide is conventionally used to form a low-resistive interface to the terminals of circuit elements such as the source and drain of MOSFETs or the terminals of resistors. It is formed by the reaction of a metal (usually Titanium) with silicon, which provides a low-resistance, thermally stable, and CMOS-compatible solution for realizing contacts. Silicide, in combination with polysilicon (or diffusion), can make resistors which are highly conductive and, as shown in Table 2.1, exhibit behavior that lies between that of poly (or diffusion) and metal. Silicided poly (or diffusion) resistors have a very low R_{sh} of $\sim$ 5 to 10 $\Omega/\square$, which is considerably higher than that of metal resistors. Their VC is lower than that of unsilicided resistors since their resistance is mainly determined by the silicide layer [2.8],[2.9]. Silicided p+ poly resistors exhibit a very low flicker noise [2.9],[2.10] and also a high TC of $\sim$ 0.3 %/°C, both similar to that of metal resistors. Their relatively high R_{sh} compared to metal resistors makes them a suitable choice for realizing TDCs. In some modern FinFet technologies, silicided resistors are not part of the standard fabrication process. Therefore, metal resistors can be employed as an alternative [2.11].

2.1.2 Readout circuits

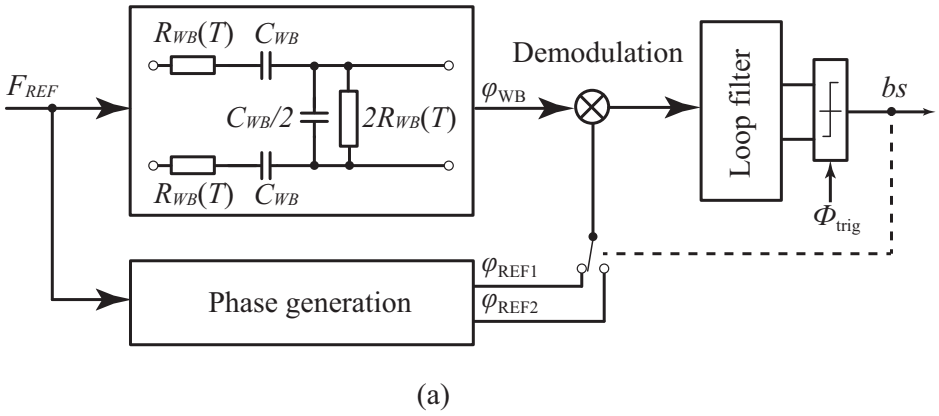
Fig. 2.2 (a) shows one method for digitizing a resistor's value that is used in a variety of resistor-based TDCs [2.9],[2.12],[2.13]. This involves digitizing the phase shift of a Wien Bridge (WB) ϕ_{WB} with respect to a reference phase ϕ_{REF}. When driven by a sine-wave signal, the output of the WB exhibits a phase-shift ϕ_{WB} with respect to the input sine-wave. This output signal is then demodulated by multiplying it by a

sine-wave with the same frequency but a different phase ϕ_{DEM}. This yields a DC term, which is a function of $\phi_{WB} - \phi_{DEM}$, along with a higher-frequency term that can be disregarded when sufficiently filtered. In practical implementations, generating a sine-wave reference signal with a precise phase reference and employing an analog multiplier is often not feasible. Instead, a square wave signal and a square-wave mltiplier, i.e., chopper switches, are utilized. It can be demonstrated that this approach only minimally reduces the DC term generated in the demodulation process. To digitize ϕ_{WB}, a phase-domain $\Sigma\Delta$ ADC can then be used to balance ϕ_{WB} with respect to two demodulation signals with phases ϕ_{REF1} and ϕ_{REF2}. When the modulator is stable, its decimated bitstream bs is a function of ϕ_{WB} and the two phase references ϕ_{REF1} and ϕ_{REF2}.

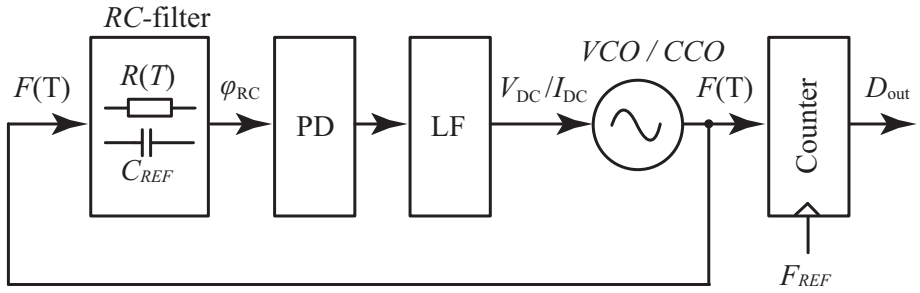
Over temperature, $\phi_{WB}(T)$ is a function of the bridge resistance $R_{WB}(T)$, the bridge capacitance $C_{WB}(T)$ and the driving frequency $F_{REF}(T)$. To be able to extract $R_{WB}(T)$ from the digitized $\phi_{WB}(T)$, it is essential that $F_{REF}(T)$ and $C_{WB}(T)$ are independent of temperature, or at least that their temperature dependence is known. On-chip Metal-Insulator-Metal (MIM) capacitors have an extremely low TC ($\sim$ 15 ppm/°C), resulting in a nearly constant C_{WB}, while a fixed $F_{REF}(T)$ can be provided externally. As a result, $\phi_{WB}(T)$ can be mainly defined by $R_{WB}(T)$. One main drawback of this method is the need for an accurate external F_{REF}, which also needs to have very low jitter (e.g., 1 ps_{rms} used in [2.9]) to avoid adding extra phase error to ϕ_{REF1} and ϕ_{REF2}.

Another method for digitizing a resistor's value is to integrate it in a Frequency-Locked-Loop (FLL) as demonstrated in [2.14],[2.15]. Fig. 2.2 (b) illustrates a simplified diagram for such a readout circuit. The FLL comprises an RC filter, a Phase-Detector (PD), a Loop Filter (LP), and either a Voltage-Controlled Oscillator (VCO) or a Current-Controlled Oscillator (CCO). The RC filter introduces a phase-shift, ϕ_{RC}, at the driving frequency, $F(T)$. This phase shift is detected and then converted into a DC voltage or current by using a PD and the LF, respectivley. The VCO or the CCO then outputs $F(T)$. At steady-state, the loop enforces a zero at the output of the PD, thereby keeping the value of ϕ_{RC} fixed. This, in turn, requires $F(T)$ to become a function of the RC components, $R(T)$ and C_{REF}. The value of $F(T)$ can then be digitized using a counter that is sampled at a much higher frequency, F_{REF}. Similar to the previously discussed readout method, when both F_{REF} and C_{REF} are held constant, the digitized value of $F(T)$ is determined by the value of $R(T)$.

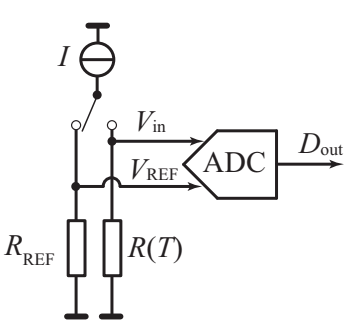
An alternative method of digitizing a resistor's value $R(T)$ is to use a reference resistor R_{REF} and then digitize their ratio using an ADC. Fig. 2.2 (c) shows an example of a readout circuit using an external R_{REF} in which an arbitrary bias current



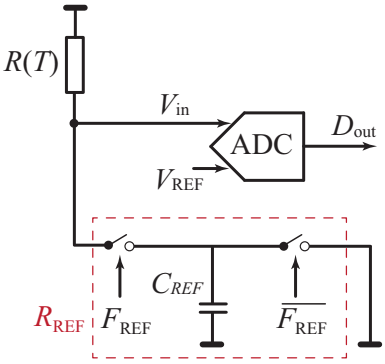
(a)



(b)



(c)



(d)

Figure 2.2: Different methods for digitizing a resistor's value: (a) using a phase-domain $\Sigma\Delta$ modulator, (b) using an FLL, (c) using an external R_{REF} and a voltage-domain ADC, and (d) replacing R_{REF} with its switched capacitor equivalent.

excites both $R(T)$ and R_{REF}. The ratio of output voltages can then be digitized by a voltage-domain ADC [2.16]. The requirement for an external component, however, rules out this method in many applications. Instead, an equivalent on-chip R_{REF} can be realized, albeit at the cost of less accuracy. As shown in Fig. 2.2 (d), R_{REF} can also be replaced by its switched-capacitor equivalent using an external reference frequency F_{REF} and an on-chip capacitor [2.17]. This results in a reference resistor equal to $1/(C_{REF} \cdot F_{REF})$, where C_{REF} is the value of the on-chip capacitor. Similar to the phase-domain readout circuit shown in Fig. 2.2 (a), this method also requires an accurate and low-jitter frequency reference [2.17].

The main drawback of the aforementioned readout circuits is that they require external references or components, which may not be readily available in many applications. However, when such limitations are not a concern, FLL-based TDCs have been shown to be a suitable option for achieving low area consumption [2.14],[2.15]. Moreover, they have exhibited decent accuracy after a single-point trim across a wide temperature range from $-40°$C to $+180°$C [2.18].

Ratio-metric readout circuits composed of *two* temperature-dependent resistors can also be used to digitized temperature. Such stand-alone resistor-based TDCs digitize a ratio of two resistor values which can then be mapped to temperature. One example of such a readout circuit, shown in Fig. 2.3, is used in [2.10] and [2.19]. For simplicity, the single-ended realization is shown in this figure. Two types of resistors are configured as a Wheatstone bridge (WhB), which provides a temperature-dependent current I_{sig}. This current is then balanced against a reference current I_{DAC} using a $\Sigma\Delta$ loop that eventually reduces the error signal $I_{error} = I_{sig} - I_{error}$ to zero. I_{DAC} is made from one of the resistor types, and the same supply voltage is used to generate the currents. As a result, the output of the TDC becomes a ratio which only depends on the value of the two resistors. Due to the combined temperature dependence of the two resistor types in this ratio, these sensors are the most energy efficient type of on-chip TDCs [2.20].

2.1.3 Temperature-sensing inaccuracy and calibration

Neglecting its voltage dependence, the value of a resistor as a function of temperature can be expressed as:

$$R(T) = R(T_r)\left(1 + TC_1 \cdot (T - T_r) + TC_2 \cdot (T - T_r)^2\right) , \tag{2.1}$$

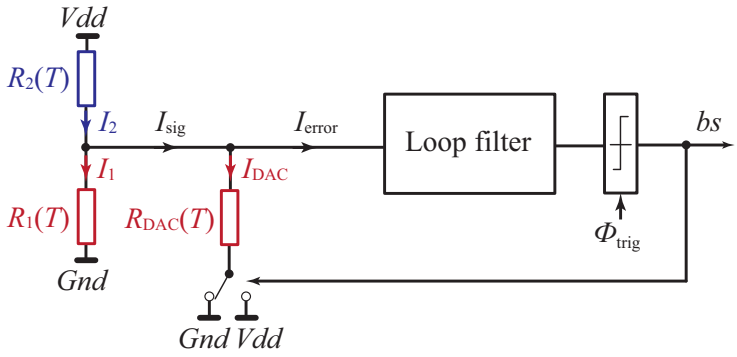


Figure 2.3: Stand-alone resistor-based readout circuit using two temperature-dependent resistors, and a ΣΔ modulator to digitize their ratio.

where $R(T_r)$ is its absolute value at a reference temperature T_r, and TC_1 and TC_2 are the first- and second-order temperature coefficients (TCs). The resistor's higher-order TCs are neglected in (2.1), which is a reasonable assumption for most temperature-sensing applications. The inaccuracy and the required calibration for resistor-based TDCs thus depend on the variations in $R(T_r)$, TC_1 and TC_2 due to processing spread and the mechanical stress caused by packaging. Unlike BJT-based TDCs, which only have to deal with one major source of spread (see Sec. 2.4), resistor-based TDCs must cope with variations of multiple parameters and thus usually require multiple calibration points (≥ 2) to achieve comparable accuracies [2.9]. This is the main disadvantage of resistor-based TDCs, which, in turn, limits their use in low-cost applications. In highly accurate MEMS frequency references, however, multiple-point calibration is often required and so resistor-based TDCs have been used to compensate for their temperature dependence [2.17]. Furthermore, resistor-based TDCs are capable of obtaining the μK-level resolution required to avoid increasing their phase noise [2.9],[2.13]. Lastly, resistors can be co-integrated with the MEMS, which can then provide good thermal tracking and thus robust compensation.

The sensitivity of a resistor's parameters to mechanical stress can also cause its accuracy to degrade after packaging. This stress sensitivity originates from either the piezo-resistive effect or from dimensional changes in the resistor due to strain. For metal resistors, this change is dominated by the dimensional variation [2.21], while for semiconductor resistors, it is dominated by the piezo-resistive effect [2.22]. Piezo-resistive coefficients, i.e., the variation of resistivity as a function of the applied stress are often empirically found [2.23],[2.24].

The piezo-resistive coefficient of a semiconductor material is a function of tem-

perature, carrier mobility and doping concentration [2.25]. It is known that lower doping concentrations result in a higher piezo-resistive coefficient [2.25]. Therefore, the stress sensitivity of n-well resistors is much higher than that of heavily doped diffusion resistors. Poly resistors are also stress-sensitive, but less than n-well or diffusion resistors [2.26]. Finally, metal resistors have the lowest stress sensitivity. Temperature also affects the piezo-resistive coefficient. It has been shown that at $-75°$C, the piezo-resistive coefficient can be $2\times$ higher than at $+175$ °C [2.25].

For a single-crystalline p-type resistor, the stress of 150 MPa creates a relative error $\Delta R/R$ of $\sim 7\%$ [2.22]. For this p-type resistor, piezo-resistive coefficients have opposite signs for different orientations of the resistor. As a result, a series combination of two resistors that are perpendicular to each other could greatly reduce the effect of stress [2.23]. For an n-type resistor, however, such cancellation found to be not perfect [2.24]. Semiconductor resistors are generally more sensitive to stress ($\sim 15\times$ to $40\times$) than vertical BJTs [2.27]. Even with *L-shaped* layout, they are at least 2-3$\times$ more stress-sensitive. The higher sensitivity of the resistors to mechanical stress imposes the need for post-packaging multi-point calibration and also introduces long-term reliability issues. The latter is because packaging stress can change over time due to variations in the encapsulation material used [2.28]–[2.30].

In [2.12], the accuracy of TDCs based on four different types of resistors (n-well, unsilicided n+ poly, unsilicided n+ and p+ diffusion resistors) implemented in a 0.18-μm CMOS technology is compared. The results show that the n+ poly resistor is the most accurate, resulting in 0.24% relative temperature-sensing inaccuracy ($\pm 3\sigma$) after a three-point calibration. Similar results are shown in [2.13] using an n+ poly resistor which achieves 0.2% relative inaccuracy. However, silicided poly resistors achieve better accuracy than unsilicided ones. In [2.9], use of a silicided p+ poly resistor in the same 0.18-μm CMOS technology resulted in 0.05% relative inaccuracy after a two-point trim. In the same work, while using the same readout circuit, an unsilicided n+ poly resistor achieved 0.48% relative inaccuracy after a two-point trim, which is about $10\times$ worse. The flicker noise corner of the silicided p+ poly resistor has also been measured to be about $10\times$ lower than that of the unsilicided n+ poly resistor: 1 Hz versus 10 Hz. The silicided p+ poly resistor also shows less stress sensitivity than the n+ poly. When the TDCs are packaged in plastic, the silicided p+ poly sensor shows a shift of about ~ 0.2 °C, while this is about ~ 1.4 °C and ~ 2.5 °C for n+ poly and highly resistive poly (undoped polysilicon) resistors. TDCs based on silicided p+ poly and n+ poly in a 65-nm technology achieve relative inaccuracies of 0.19% and 0.34% after a two-point trim [2.31]. Using a one-point trim, the inaccuracy achieved is about $20\times$ worse: 4% and 7.2%.

Stand-alone TDCs using two types of resistors are expected to have worse accuracy and stress sensitivity than those that use only one type. This is due to the combined effect of spread and stress on the two types of resistors, which is unlikely to be correlated or canceled out. In [2.19], the combination of an n+ poly and a p+ diffusion is used, resulting in a relative inaccuracy of 1.1% (Max-Min) after a one-point trim. This then drops to 0.23% after a two-point trim. A better result is obtained by combining a silicided p+ poly and an n+ poly resistor, as done in [2.10]. Although the TDC shows a large untrimmed inaccuracy of 8.3%, using a one-point trim can reduce this to 1.1% (3σ), and using a two-point trim can reduce it even further to 0.13%. Slightly better accuracy has been demonstrated in [2.20] using a silicided diffusion resistor combined with an n+ poly resistor.

2.2 Thermal diffusivity (TD) based TDCs

2.2.1 Principle of operation

Another method for realizing a CMOS TDC is to measure the temperature-dependent rate at which heat diffuses through silicon [2.32],[2.33]. The thermal diffusivity of silicon D_{si}, is a measure of the speed at which heat diffuses through silicon. $D_{si} = k_{si}/(\rho_{si} \cdot C_{si})$, where k_{si} is the heat conductivity, ρ_{si} is the density of silicon and C_{si} is its thermal capacity.

In standard CMOS, the substrate is made from lightly doped silicon, and so its thermal characteristics are close to that of pure silicon. As a result, D_{sub} is similar to D_{si}, and it is reasonably process-independent. D_{si} is also a strong function of absolute temperature [2.32]. Over the industrial temperature range, D_{si} can be approximated to within a few percentage points with a power law, i.e., $D_{si} \propto T^{-n}$, where the n of pure silicon is ~ 1.8 [2.34]. Measurements show that D_{si} varies from 2 cm^2s^{-1} at 200 K to 0.53 cm^2s^{-1} at 400 K [2.34].

2.2.2 Readout circuits

In order to measure D_{sub}, thermal-diffusivity (TD) sensors based on electro-thermal filters (ETFs) are used. An ETF consists of a heater and a relative temperature sensor placed in close proximity (see Fig. 2.4). An ETF behaves like a low pass filter in thermal domain, exhibiting a phase shift between the output of the temperature sensors and the applied heat [2.35]. This phase shift is not only a function of the frequency of the applied heat but also of the geometry of the ETF and D_{sub} [2.33].

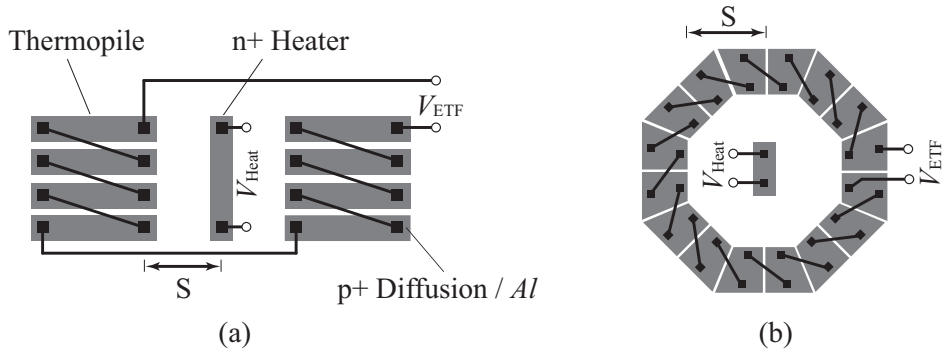


Figure 2.4: Electro-thermal filters (ETF) realized in CMOS. (a) A bar-type ETF, taken from [2.33]. (b) optimized ETF by placing a thermopile in a fixed phase shift contour around the heater, taken from [2.38].

The geometry is defined by lithography, which can be considered very accurate for relatively large ETF dimensions. The phase shift of the ETF at a known frequency will thus be mainly determined by D_{sub}.

The objective of the readout circuit of a TD sensor is to digitize the phase shift of the ETF. This can be done by incorporating it in a Frequency-Locked Loop (FLL), which generates a temperature-dependent output frequency, as shown in Fig. 2.5 [2.32],[2.33]. The output frequency can be digitized using a higher frequency reference provided externally [2.33],[2.36]. Alternatively, the phase shift of an ETF driven by a known frequency, F_{REF}, can be digitized with a phase-domain $\Sigma\Delta$ modulator similar to that used in the resistor-based TDCs shown in Fig. 2.2 (a), except that the ETF replaces the WB [2.36],[2.37].

As shown in Fig. 2.4 (a), the heater of the ETF is realized with n+ diffusion, while the temperature sensor consists of thermocouple junctions of p+ diffusion and *Al* that are placed at a distance s from the heater's centroid. To further maximize the sensitivity, a number of thermocouples are daisy-chained around the heater to form a thermopile. The thermopile is a relative temperature sensor which creates an output voltage proportional to the Seebeck coefficient of the junction and the temperature difference between the hot and cold junctions at each side of the thermocouple arm. Taking the temperature gradient of the heat pulses into account, reducing s causes a larger signal between the hot and cold junctions, but also means that the distance s will become comparable to the ETF's geometry, resulting in lower accuracy [2.33]. This results in a trade-off between accuracy and resolution in TD-based sensors.

One main consideration in designing TD-based sensors is the mitigation of additive

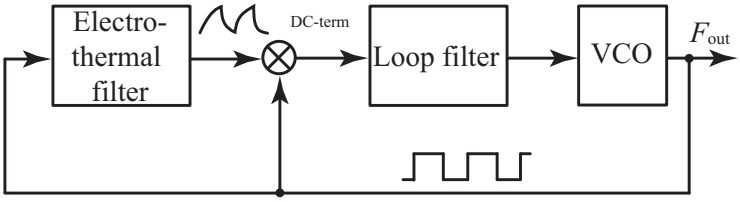


Figure 2.5: Thermal-diffusivity based sensor using a frequency-locked-loop (FLL) readout circuit. The output frequnecy is proportional to the phase shift of the ETF.

phase shift errors. These, for instance, can originate from the electrical filtering caused by the resistance of the p+ arms of the thermocouple with the capacitance formed between them and the n-well. Similarly, an additional phase shift can be caused by any readout circuitry that is directly connected to the ETF's output, for instance, an amplifier. Although increasing the arm length of the ETF increases the hot-cold difference and results in a higher signal, it also increases the resistance, electrical phase shift, and thermal noise. When it is not calibrated, the large spread of thermocouple resistance can contribute substantially to the temperature-sensing errors [28]. Since these resistances are also temperature-dependent, they can create a temperature-dependent error that cannot be removed with a one-time calibration. Therefore phase calibration might be required, as suggested in [2.38]. Another major drawback of TD-based sensors is their relatively large power consumption (a few mW) most of which ($> 80\%$) is consumed by the heater to create a sufficiently large electrical signal at the output of the thermopile ($\sim 1 - 5$ mV). Compared to other types of TDCs, TD-sensors consume the most power.

2.2.3 Temperature-sensing inaccuracy and calibration

Without trimming, TD-based sensors in 0.7-μm CMOS technology achieve relative temperature-sensing inaccuracies of 0.69% (3σ) and 1.38% (3σ), respectively, for thermopiles with $s = 20$ and 10 μm [2.36]. By optimizing the layout of the ETF in the same technology, the inaccuracy achieved could be improved by $2\times$ [2.37]. Compared to the parallel placement of thermocouple arms on each side of the bar-type heater in [2.36], in the optimized ETF in [2.37], the arms are placed in a fixed phase-shift

contour around a central heater. Fig. 2.4 (b) shows an example of an optimized ETF in which the resistance of the thermopile is also reduced [2.38]. The improved inaccuracy in [2.37] is mainly due to the large choice of s (= 24 μm) compared to the ETF's geometry, which reduced the SNR. This, in turn, means that a relatively long conversion time of $\sim$ 6.25 s is required to obtain a 30 mK_{rms} [2.37]. Using the same ETF dimensions of [2.37], but realized in a 0.18-μm technology, results in an improved inaccuracy due to the higher accuracy of the lithography, with a 0.22% inaccuracy [2.39].

TD-based sensors can also be made very small, occupying for instance 0.0017 mm^2 in a 40-nm standard CMOS technology [2.38], which is one of the smallest temperature sensors reported to date. This level of area consumption is mainly achieved due to better lithographic accuracy allowing both s to be reduced, and the use of mostly-digital phase-domain readout circuitry [2.38]. The distance s (= 2 μm) chosen in this TD sensor is $\sim 10\times$ lower than the previously mentioned TD sensors. The sensor is also realized in an octagonal thermopile shape (see Fig. 2.4 (b)), which increases it's SNR. After a batch-calibration to extract the average parameters of a fifth-order master curve, the TD sensor in [2.38] achieves inaccuracies of 1.7% (3σ) and 2.79% (3σ) for ETFs with $s = 3.3$ μm and $s = 2$ μm, respectively. Compared to [2.37] and [2.39], the TD-based sensor in [2.38] also achieves a substantially larger bandwidth of 500 Hz, which is an improvement of more than $\sim 1000\times$. Increasing the driving frequency, however, imposes an additional electrical phase shift on both the ETF and the readout circuit, resulting in the need for a background phase calibration to remove them from the phase shift of interest. After a continuous phase calibration, the inaccuracy of the two TD-based sensors drops to 0.6% (3σ) and 1% (3σ), respectively.

Based on the implemented TD-based sensors in standard CMOS technology nodes from 0.7-μm to 40-nm CMOS, it can be concluded that TD sensors can be very accurate when the ETF is realized with a relatively large s ($\sim$ 10 μm): untrimmed inaccuracies of $\sim \pm 0.2$°C (σ) from -55°C to $+125$°C [2.40]. The ability to achieve such decent accuracy without trimming is the main advantage of TD sensors. In [2.41], this specific characteristic of TD sensors is exploited to self-calibrate a low-power and energy-efficient resistor-based sensor. Since both sensor types can be read out by a phase-domain ADC, the area overhead is minimal. Ultimately, the TD sensor, achieving a 3σ-inaccuracy of $\sim$0.2°C, is employed to self-calibrate the resistive sensor at two temperatures, significantly reducing its 3σ-inaccuracy from $\sim$10°C to $\sim$0.25°C.

In [2.40], however, TD sensors exhibited a batch-to-batch shift of about 2°C. Although this could be corrected by a single-point trim, this removes most of the accuracy advantages of TD sensors. One hypothesis is that this shift is caused by the

spread of the doping in the substrate and the spread of thickness in epi-layer [2.40], which changes the thermal properties of ETF. In line with this, sensors fabricated in an SOI process exhibited 10× less batch-to-batch shift [2.40]. The effect of packaging on TD-based sensors has been reported in [2.38], [2.39] and [2.40], which show an increase in the sample-to-sample spread with plastic packaging compared to that with relatively stress-free ceramic packaging. This is due to either the sensitivity of D_{si} to mechanical stress [2.42], or to the extra self-heating in the plastic packaging [2.40]. However, the shift is only about ±0.1°C [2.40], which eliminates the need for post-package trimming for TD sensors.

2.3 MOSFET-based TDCs

2.3.1 Principle of operation

The temperature-dependent characteristic of MOSFETs can also be used to realize TDCs. The key advantage of MOSFET-based TDCs is their ability to operate with low supply voltages (e.g., 0.5 V), dissipating very little power (e.g., tens of nanowatts), or ocuppying a very small area (e.g., 0.001 mm^2) [2.43]–[2.45]. However, they achieve poor accuracies due to the several process dependent parameters that determine MOSFET behavior, and they are sensitive to packaging stress. As a result, MOSFET-based TDCs require a multi-point (post-packaging) calibration to obtain accuracies comparable to those of BJT-based TDCs. When the operating temperature range is limited, for instance in health applications, or when ultra-low power consumption is required, for instance in IoT sensing nodes, MOSFET-based TDCs are suitable candidates [2.43],[2.46].

MOSFETs can be biased in different operating regions. When the Gate-Source voltage V_{GS} surpasses the threshold voltage V_T, the channel under the gate is inverted, i.e., the concentration of mobile carriers at the surface (with opposite polarity) exceeds the dopant density in the bulk. In this inversion region, depending on the Drain-Source voltage V_{DS}, the MOSFET operates in either the triode or saturation region. The drain current then becomes a function of V_{GS}, V_{DS}, mobility μ, and V_T.

When V_{GS} is below V_T, the channel under the gate of a MOSFET is not inverted. Nevertheless, the number of mobile carriers at the surface exhibits an exponential dependence on V_{GS}. In this so-called sub-threshold region, applying a V_{DS} that is larger than the drain saturation voltage V_{DSAT} results in a current that can be

approximated as:

$$I_{D,sth} = (m-1)\,\mu C_{ox}\frac{W}{L}\left(\frac{kT}{q}\right) exp\left(\frac{q\cdot(V_{GS}-V_T)}{mkT}\right)\cdot\nu_{sat}\,, \tag{2.2}$$

where m is the subthreshold factor defined as $m = 1 + C_D/C_{ox}$ that can range from 1.1 to 1.4, C_D and C_{ox} are the depletion and oxide capacitances, and ν_{sat} is the saturated velocity. This exponential $I_D - V_{GS}$ relationship resembles the $I_C - V_{BE}$ characteristics of BJT, albeit with additional parameters involved.

A MOSFET can also be used as a dynamic threshold device: DTMOST [2.47]–[2.49]. In this configuration, the gate and the well of a MOSFET are tied together as shown in Fig. 2.6. This prevent the formation of a strong channel under the gate, and thus, the device exhibits behavior similar to a subthreshold MOSFET. A DTMOST can also be viewed as a lateral PNP device with an extra polysilicon gate placed on top of the base region resulting in a lower *apparent bandgap voltage* [2.47]. This is due to a built-in voltage appearing between the polysilicon gate and the silicon well that will then be divided between the C_{ox} and C_D of the MOSFET. In a 0.35-μm technology, this is ~ 0.6 V, which is about half of the silicon bandgap voltage.

The temperature dependence of MOSFETs relies on that of the parameters involved in their model. Among the most critical parameters are μ and V_T, whose temperature dependence is often described by simple empirical model. As discussed in A.3, $\mu(T) = (T_r)(T/T_r)^m$, where m is a process-dependent constant, and $\mu(T_r)$ is the value of μ at a reference temperature T_r. Similarly, the threshold voltage is given by $V_T(T) = V_T(T_r) - \alpha_{V_T}(T - T_r)$, where α_{V_T} is a process-dependent constant ranging from 1 to 4 mV/°C, and $V_T(T_r)$ is the value of V_T at a reference temperature T_r. To understand the variations of the parameters m and α due to process or stress, a thorough understanding of the physical dependence of μ and V_T is necessary. For instance, V_T relies on several physical parameters, including intrinsic carrier concentration, dopant density, oxide capacitance per unit area, permittivity of silicon, metal-semiconductor work function difference, and fixed charges in the oxide layer.

2.3.2 Readout circuits

MOSFET-based TDCs have been realized with different readout circuits. These circuits can best be classified in three main categories, which operate based on the digitization of the (1) temperature-dependent propagation delay of a delay cell, (2) temperature-dependent frequency of an oscillator or a ring oscillator, and (3) temperature-dependent ratios of voltages or currents derived from MOSFET charac-

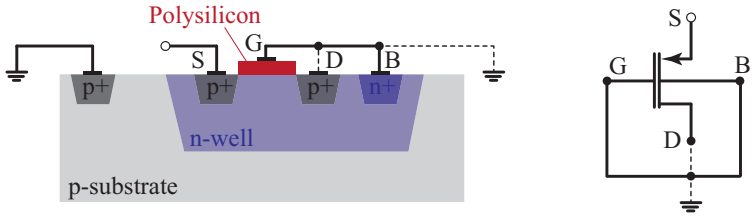


Figure 2.6: P-type dynamic threshold MOS transistor (DTMOST) realized in a CMOS technology.

teristics.

In the first category, the readout circuit is based on digitizing the width of a pulse, which is determined by the propagation delay (t_{pd}) of some delay cells [2.43],[2.50]. The pulse-width is controlled by controlling its rising edge, falling edge, or both. This pulse-width is then digitized by using a counter clocked by a higher frequency reference. In [2.50], the pulse-width is determined by the propagation delay of CMOS inverters that control only one edge of the pulse. t_{pd} is thus related to dimensions, V_T and μ of the MOSFETs in the inverter and also the supply voltage. In [2.43], the pulse-width is defined by the t_{pd} of two delay cells, which independently determine the rising and falling edges of a pulse, as shown in Fig. 2.7. Unlike using an inverter as the delay determining element, the delay of these two cells is controlled by two temperature-dependent currents derived from MOSFETs in the subthreshold region. V_{GS} and V_{DS} of the MOSFETs are converted into current using V-I converters. Several other examples have been shown in the literature for the digitization of temperature-dependent delays, for instance in [2.51] and [2.52]. One drawback of these delay-based readout circuits is the dependency of t_{pd} to supply variations, which can be as high as 1 °C/mV [2.53]. The challenge is, then, to provide a regulated supply voltage for the delay cells which is also temperature-independent.

The second category of readout circuits is based on the digitization of the temperature dependent frequency of an oscillator (or two oscillators). This process is very similar to that of delay-based readout circuits, as the frequency of oscillation is also defined by some propagation delays. Fig. 2.8 illustrates one example, in which I_1 and I_2 are derived from the temperature-dependent characteristics of MOSFETs that determine the frequencies of two ring oscillators. The ratio of these frequencies is then obtained by using an asynchronous counter as shown in Fig. 2.8 [2.54],[2.55]. A temperature-dependent frequency can also be obtained by using temperature-dependent delay cells inside the ring oscillator as used in [2.56]–[2.58]. This category of MOSFET-based

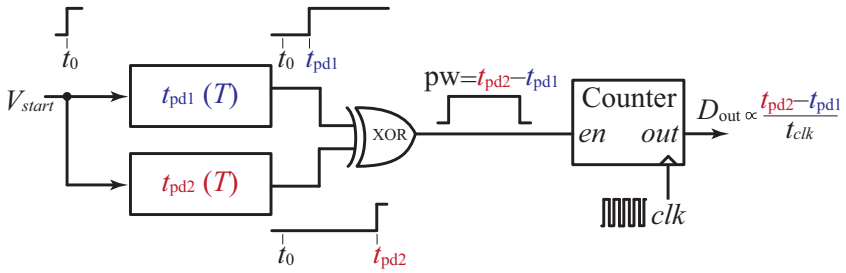


Figure 2.7: Readout circuit using temperature-dependent propagation delays.

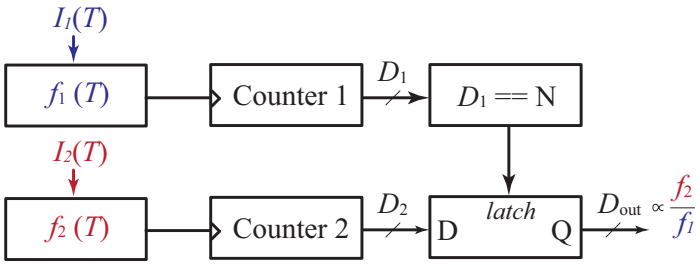


Figure 2.8: Readout circuit using temperature-dependent frequencies.

TDCs has also showed a high level of supply dependence. Mitigating this sensitivity thus remains an ongoing challenge, with efforts focused on employing circuit techniques and implementing effective supply regulation strategies [2.56],[2.57],[2.59].

In the third category of readout circuits, a voltage-domain ADC is used to digitize the ratio of two temperature-dependent voltages derived from MOSFETs. Similarly, this can be done by obtaining temperature-dependent currents or charges and then digitizing their ratio with the corresponding current or charge domain ADCs. In [2.48],[2.49], temperature-dependent signals are derived from two DTMOSTs that are biased at a fixed current ratio as illustrated in Fig. 2.9. The V_{GS} (V_{GS2}) of a DTMOST has a CTAT characteristic, while the difference between the two V_{GS}, i.e., $\Delta V_{GS}(= V_{GS2} - V_{GS1})$, has a PTAT characteristic. The combination of V_{GS} and ΔV_{GS} can result in a flat reference voltage V_{REF} $(= V_{GS} + \alpha \cdot \Delta V_{GS})$, where α is a gain factor. A voltage-domain ADC can then digitize the ratio $\alpha \cdot \Delta V_{GS}/V_{REF}$ as an indication of temperature. Combining V_{GS} and ΔV_{GS} has also been used to realize low-voltage (∼0.6 V) bandgap voltage references [2.47].

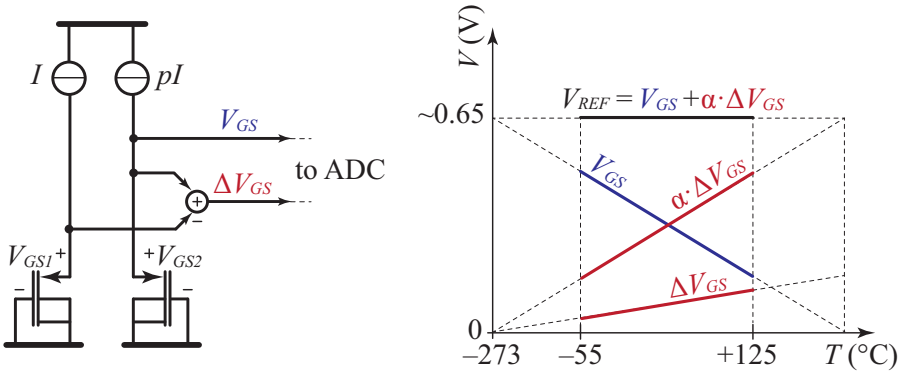


Figure 2.9: Operation principle of DTMOST-based temperature sensors, taken from [2.48].

2.3.3 Temperature-sensing inaccuracy and calibration

Based on the reported results in the literature [2.60], MOSFETs in the saturation and triode regions are usually less accurate than MOSFETs in either the subthreshold region or configured as a DTMOST. As an example, in [2.61] a current is derived from a MOSFET in its triode region, which is then used to obtain temperature-dependent delay. Realized in a 0.18-μm CMOS technology, and using a time-domain readout circuit, the TDC achieves a relative inaccuracy of 2% (Max-Min) after a two-point calibration. The TDC in [2.53] is another example in which delay cells of a ring-oscillator are determined by MOSFETs in their saturation region. Realized in 65-nm CMOS technology, this TDC achieved inaccuracies of 2.73% (Max-Min) using one-point calibration.

In the subthreshold region, MOSFET-based TDCs seem to be more accurate. The TDC in [2.54], which is realized in a 0.18-μm CMOS technology, uses two MOSFETs in the subthreshold region to generate two temperature-dependent currents. These MOSFETs have the same dimensions and V_{GS} but different V_{DS}. The ratio of the two currents is then digitized to generate a function of temperature, achieving temperature-sensing inaccuracies of 2% (Max-Min) after a two-point calibration. Similar results have been obtained in 65-nm CMOS technology by using two MOSFETs in the subthreshold region [2.51],[2.55]. Unlike [2.54] and [2.55], the MOSFETs in [2.51] have the same V_{DS} but different V_{GS} for generating the temperature-dependent currents. It then achieves an inaccuracy of 3% (Max-Min) after a two-point calibration. Better results have been reported in [2.56], in which MOSFETs in the subthreshold

region are used as the delay cell of a ring oscillator obtaining 0.42% relative inaccuracy after a two-point calibration. It is worth mentioning that the TDCs in [2.51], [2.54] and [2.56] all consume less than 1 μW, while [2.51] and [2.54] could also operate below 0.8 V supply voltages.

MOSFETs configured as DTMOSTs have shown the best in class temperature-sensing inaccuracies and supply sensitivity. In a 0.16-μm CMOS technology, DTMOST-based TDCs achieve inaccuracies of 0.48% (3σ) after a one-point calibration [2.48],[2.49]. In 28nm CMOS, a sensor achieves inaccuracy of 1.8% (3σ) after a one-point calibration [2.62], while obtaining the lowest supply sensitivity among MOSFET-based TDCs.

One drawback of MOSFET-based TDCs is the higher sensitivity of MOSFETs to mechanical stress compared to BJTs. This not only results in costly post package calibration but may also cause long-term stability issues. It is known that mobility increases with the applied stress, e.g., 200 MPa stress causes $\sim$ 10% variation in mobility [2.63]. The threshold voltage of long channel MOSFETs, on the other hand, is at least $\sim 10\times$ less sensitive to stress than mobility [2.64], and can therefore be ignored. In short-channel MOSFETs, the drain saturation voltage V_{dsat} is affected by the velocity saturation, making V_{dsat} more sensitive to stress than the V_T of long channel devices. The V_{dsat} of an NMOSFET varies by $\sim -1.5\%$ and $\sim -3.5\%$, for $\sim$ 140 MPa tensile and compressive stress, respectively [2.64].

2.4 BJT-based TDCs

2.4.1 BJTs in CMOS Technology

The most common method for realizing a TDC is to use Bipolar Junction Transistors (BJTs). In standard CMOS processes, two types of BJTs can be realized: vertical and lateral.

In vertical BJTs, the direction of current flow is perpendicular to the silicon surface. In the case of vertical PNPs, a $p+$ diffusion functions as the emitter, the n-well is the base and the p-substrate is the collector (see Fig. 2.10 (a)). A vertical PNP can be realized in any CMOS technology, since realizing an n-well is always possible. A vertical NPN, on the other hand, can only be realized in twin-well CMOS technologies, which allow the realization of a deep n-well (see Fig. 2.10 (b)). The key disadvantage of vertical PNPs is that their collector has a low-resistive path to the p-substrate and thus inevitably has the same (ground) potential. As a result, the only accessible terminals are the emitter and the base, which can then be used to

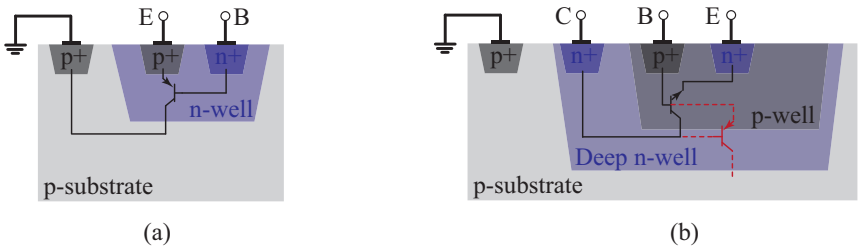


Figure 2.10: Vertical BJTs in CMOS technologies: (a) Vertical PNP and (b) vertical NPN. Red shows the parasitic vertical PNP associated with the vertical NPN.

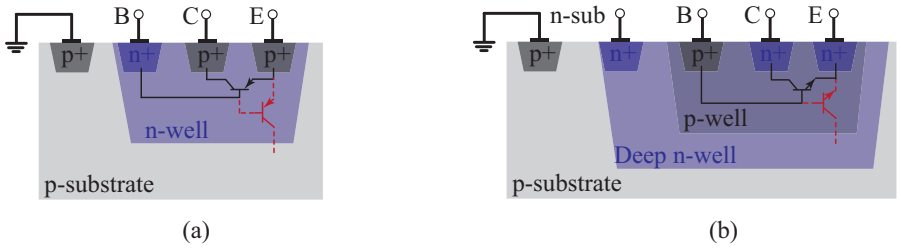


Figure 2.11: Lateral BJTs available in CMOS technologies. (a) Lateral PNP and (b) lateral NPN. Red shows the parasitic vertical devices.

bias the device. This is not the case for vertical NPNs, as the p-well potential can be freely chosen. Vertical NPNs, however, come with a parasitic vertical PNP with a grounded collector, as shown in Fig. 2.10 (b). This causes the V_{BC} of the main NPN to appear as the V_{BE} of the PNP device, whose emitter current will be added to the base current of the NPN. One way to avoid this current is to use the vertical NPN in a diode-connected configuration (i.e., its $V_{BC} = 0$), in which the PNP has zero potential across its V_{BE}.

Lateral BJTs are made using doped regions near the surface that are usually used to form the drain and source of MOSFETs. In the case of a lateral PNP, $p+$ regions can form the collector and emitter terminals while an n-well forms the base. Unlike the vertical PNP, a lateral PNP has a free collector that can be set at arbitrary voltages. Although this is a potential advantage, this device is used less frequently compared to the vertical PNP. One reason is that there is an extra vertical device (see Fig. 2.11 (a)) that adds to its emitter current. In conventional bipolar technology, a buried n+ region prevents the formation of this vertical PNP. Similarly, in twin-well CMOS technologies, the deep n-well can be used to prevent currents from flowing through the vertical emitter (see Fig. 2.11 (b)). In standard CMOS technology, this layer does not

exist, and therefore lateral PNPs are inevitably associated with the vertical device. To reduce the effect of the parasitic device, the ratio I_{CL}/I_{CV} must be maximized, where I_{CL} is the lateral collector current, and I_{CV} is the vertical collector current. This can be achieved by minimizing the emitter area and the lateral base width, and by having the collector surround the emitter [2.65]. Experimental results show that the ratio I_{CL}/I_{CV} can be 1.5 to 4, which means that the lateral bipolar contributes 60% to 80% of the total current [2.65],[2.66].

2.4.2 Principle of operation

Ideally, BJTs have an exponential $I_C - V_{BE}$ characteristic that can be expressed as:

$$V_{BE} = n_F \frac{kT}{q} ln\left(\frac{I_C}{I_S} + 1\right), \tag{2.3}$$

where V_{BE} is the base-emitter voltage, $n_F \sim 1$ is the non-ideality factor of the PNPs, I_C is the collector current and I_S is the saturation current of the device. Due to the temperature dependence of I_S, V_{BE} has a CTAT characteristic which changes by ~ -2 mV/°C. The temperature dependence of V_{BE} and the underlying physics behind I_S will be further discussed in Chapter 3. For now, it should be noted that I_S is proportional to the emitter area A_E of the device. Due to the extremely low values of I_S, the assumption $I_C/I_S + 1 \sim I_C/I_S$ is usually made in (2.3). Taking this assumption into account, when two BJTs have a fixed ratio between their collector current densities p, either due to different I_C ratios (see Fig. 2.12) or different A_E ratios, the difference between their V_{BE} becomes a PTAT voltage that can be expressed as:

$$\Delta V_{BE} = n_F \frac{kT}{q} ln\left(p \cdot \frac{I_C}{I_S}\right) - n_F \frac{kT}{q} ln\left(\frac{I_C}{I_S}\right) = n_F \frac{kT}{q} ln(p), \tag{2.4}$$

where p is the collector current density ratio of the two BJTs. The temperature dependence of V_{BE} will then depend on p and some physical constant. A key observation of ΔV_{BE} is that it barely depends on absolute parameters such as I_C or I_S. The ratio p, on the other hand, can be made accurate by circuit techniques. As a result, ΔV_{BE} becomes robust to process variations and also stress of the packaging. The two voltages with the opposite temperature dependence, V_{BE} and ΔV_{BE}, can be combined to obtain an almost flat voltage reference V_{REF} $(= V_{BE} + \alpha \cdot \Delta V_{BE})$, where α is a fixed gain factor (see Fig. 2.12). To realize a TDC, it is therefore required to digitize μ_T as the ratio of $\alpha \cdot \Delta V_{BE}$ and V_{REF}, which can then be mapped to temperature in the digital domain. In the ratio μ_T, the non-ideality factor of the

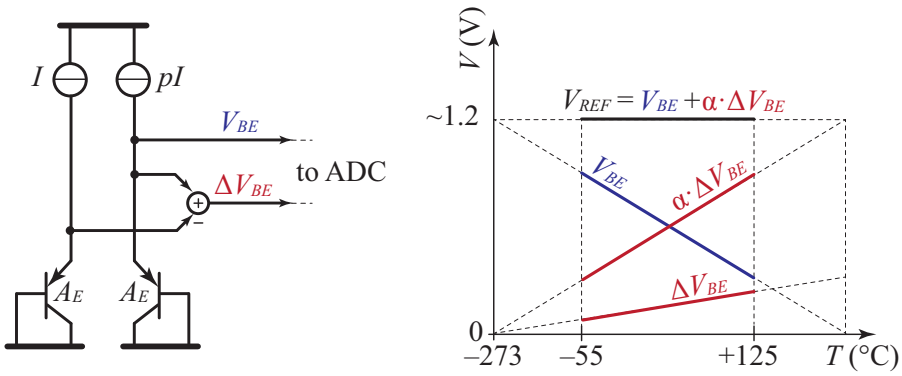


Figure 2.12: Operation principle of BJT-based temperature sensors.

PNPs n_F cancels out, since it is a multiplicative term in both 2.3 and 2.4. However, when the absolute values of V_{BE} and ΔV_{BE} are important, for instance in a bandgap voltage reference, or when performing voltage calibration (see Sec. 4.6.4), the value and spread of n_F must be taken into account.

The BJT-based sensing principle discussed so far operate under the condition of biasing a BJT with a fixed current source. These approaches typically require supply voltages that exceed 1 V because, at low temperatures, $V_{BE}(T)$ approaches 0.8 V. In recent years, due to technology scaling, there has been growing interest in BJT-based sensors capable of reliable operation with sub-1V supply voltages [2.67]. To overcome this limitation, new sensor designs have emerged that do not rely on biasing a BJT with a current source. Instead, they utilize capacitors [2.68]–[2.70].

Figure 2.13 illustrates the operation of such a capacitively-biased sensor. Initially, a capacitor is charged to the supply voltage in a reset state and then discharged through either a diode (Fig. 2.13-a) or a BJT (Fig. 2.13-b). As shown in Figure 2.13-c, the voltage on the capacitor gradually decays, following a near-ideal logarithmic function after an initial phase, as described in [2.71].

In this logarithmic region, the capacitor's voltage V_C remains almost independent of the initial phase and can be approximated as:

$$V_C(t) = n_F V_T \cdot ln(\frac{n_F V_T \cdot C}{I_S \cdot t}), \tag{2.5}$$

where V_T is the thermal voltage, given by kT/q. When V_C is sampled at a specific instant, the resulting voltage demonstrates a CTAT characteristic similar to that of

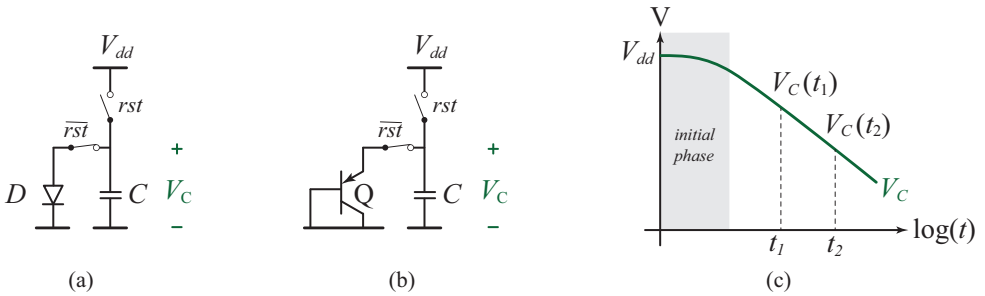


Figure 2.13: (a) A sensing frontend for capacitively-biased diodes. Storing a charge proportional to (b) V_{BE} and (c) ΔV_{BE}.

V_{BE}. This similarity can also be observed by comparing (2.5) to (2.3). When V_C is sampled at two instances, t_1 and $t_2 = n \cdot t_1$, the voltage difference between them demonstrates a PTAT characteristic, similar to ΔV_{BE}, and can be expressed as:

$$\Delta V_C = V_C(t_1) - V_C(t_2) = n_F \frac{kT}{q} \cdot ln(n)\,. \tag{2.6}$$

The ratio between V_{BE} and ΔV_{BE} can then be utilized for temperature sensing. The charge stored on the capacitor C is transferred to a voltage-domain discrete-time ADC to digitize their ratio [2.69],[2.70]. A sensor based on this approach achieved a 3σ accuracy of (± 0.15°C) from -55°C to $+125$°C after a one-point trim [2.70]. This accuracy level is quite comparable to that achieved by traditional readout circuits.

2.4.3 Readout circuits

BJT-based readout circuits can be categorized in the same way as those used in MOSFET-based TDCs. As described in section 2.4.1, the ratio between $\alpha \cdot \Delta V_{BE}$ and V_{REF} can be digitized in the voltage domain [2.72]–[2.75], or in the current domain [2.76]. Furthermore, this ratio can also be digitized in the time-domain [2.77]–[2.80], and in the frequency-domain [2.81].

Switched-capacitor readout circuits in the voltage domain have traditionally been favored for their robustness, immunity to clock jitter, and ease of integration with dynamic techniques essential for mitigating the effects of mismatch and $1/f$ noise [2.72]–[2.75]. A more comprehensive exploration of these readout circuits will be presented in Chapters 4 and 6. Nevertheless, it should be noted that these readout circuits introduce kT/C noise, which limits resolution, and energy efficiency [2.82]. Continuous-time readout circuits can circumvent this limitation, achieving higher

energy efficiency [2.76],[2.83].

When low area or power consumption is the main focus for a temperature sensor, SAR ADCs have been used [2.68],[2.84]–[2.87]. However, specific errors such as non-monotonicity or missing codes limit their application. Time-domain readout circuits also achieve low area and power consumptions [2.88],[2.89], but these sensors require external clock frequency, which limits their use.

2.4.4 Temperature-sensing inaccuracy and calibration

In CMOS technology, BJT-based TDCs achieve the highest accuracy. This is mainly because ΔV_{BE} is, to first-order, independent of absolute parameters. Obtaining an accurate TDC, thus, boils down to correcting the errors associated with the spread of V_{BE}. As will be discussed in Chapter 3, variations in V_{BE} can be effectively corrected by a one-point trim, resulting in a V_{BE} that is accurate to within $\sim \pm 200 - 300$ μV over a large temperature range. Another advantage of BJT-based TDCs is their very low stress sensitivity, which will be further discussed in Sec. 3.7. They have also shown extremely high long-term stability, for instance, a drift of ~ 6 mK after ~ 1 year [2.78].

Over the temperature range from -55°C to $+125$°C, TDCs utilizing PNPs and switched-capacitor readout circuits have exhibited relative inaccuracies (3σ) ranging from 0.56% to 0.07% after a one-point trim [2.72]–[2.75]. Expanding the temperature range up to 180°C has also proven feasible with slightly greater inaccuracy, e.g., relative inaccuracies of 0.18% [2.90], and 0.39% [2.91]. Integration of a continuous-time readout circuit and its combination with dynamic techniques has resulted in a relative inaccuracy of 0.11%, a performance comparable with TDCs employing switched-capacitor readout circuits [2.83].

Capacitively-biased sensors, combined with a switched-capacitor readout circuit result in an inaccuracy of 0.67% after a one-point trim [2.69]. Capacitively-biased sensors have demonstrated even better accuracy when PNPs were used instead of diodes, achieving a 3σ-inaccuracy of 0.17% [2.70].

In more advanced FinFet technology with feature sizes below 20nm, BJTs exhibit degraded performance [2.67],[2.92]. The larger base-width resulting from their specific realization in such technologies leads to very low current gain and non-exponential I-V characteristics. This has prompted the use of bulk diodes, which spread less than BJTs in some recent technology nodes [2.68]. The use of BJTs in the most advanced technology nodes remains a subject that requires further research and development,

but in 4nm FinFet, a sensor has already demonstrated a relative inaccuracy of 0.33% after a two-point trim [2.87].

2.5 Conclusion

In CMOS technology, TDCs can be realized based on resistors, MOSFETs, thermal-diffusivity, and BJTs. Their operating principles and their limitations in terms of accuracy and required calibration have been discussed in this chapter. The most accurate TDCs are usually based on BJTs, which obtain excellent accuracy (down to ± 0.2 °C (3σ)) with only a one-point trim. Low sensitivity to stress, long-term stability, and a very well-known temperature dependency are other reasons that these TDCs are very popular. Resistor-based TDCs based on silicided p+ poly resistors have shown comparable accuracy, but require a two-point trim. However, when they are combined with a second resistor type to realize stand-alone TDCs, the obtained accuracy is reduced. Due to the higher stress-sensitivity of resistors compared to BJTs, it is also expected that plastic-packaged TDCs will require post-packaging trim. MOSFET-based TDCs are promising candidates for realizing low-power TDCs, however, they also require multi-point calibration ($>$ 2-points) to approach the accuracy of BJT-based TDCs. TD-based sensors have shown promising accuracies ($\sim \pm 0.2$ °C (3σ)) without calibration. However, they currently exhibit significant batch-to-batch shift ~ 2°C. Since state-of-the-art BJT-based TDCs suffer from the same batch-to-batch shifts as TD sensors, they can be considered equal in terms of accuracy, and calibration requirement. However, solving their batch-to-batch shift, makes BJT-based TDCs more favorable, since they consume lower power.

Bibliography

[2.1] P. Srinivasan, W. Xiong, and S. Zhao, "Low-frequency noise in integrated n-well resistors," *IEEE Electron Device Letters*, vol. 31, no. 12, pp. 1476–1478, 2010.

[2.2] R. A. Ashton, "Measurement of n-well sheet resistance under p+ diffusion and p channel gate," in *ICMTS 1999. Proceedings of 1999 International Conference on Microelectronic Test Structures*, March 1999, pp. 45–50.

[2.3] H. M. Chuang, K. B. Thei, S. F. Tsai, and W. C. Liu, "Temperature-dependent characteristics of polysilicon and diffused resistors," *IEEE Transactions on Electron Devices*, vol. 50, no. 5, pp. 1413–1415, May 2003.

[2.4] A. Ito, "Modeling of voltage-dependent diffused resistors," *IEEE Transactions on Electron Devices*, vol. 44, no. 12, pp. 2300–2302, 1997.

[2.5] R. Brederlow, W. Weber, C. Dahl, D. Schmitt-Landsiedel, and R. Thewes, "Low-frequency noise of integrated polysilicon resistors," *IEEE Transactions on Electron Devices*, vol. 48, no. 6, pp. 1180–1187, 2001.

[2.6] Y. Tsividis, *Mixed Analog-Digital VLSI Devices and Technology*, ser. EngineeringPro collection. World Scientific, 2002.

[2.7] N. Alvi, S. Tang, R. Kwor, and M. Fulcher, "Recrystallization by rapid thermal annealing of implanted low-pressure chemical-vapor-deposited amorphous Si films," *Journal of Applied Physics*, vol. 62, pp. 4878 – 4883, 01 1988.

[2.8] E. Vereshchagina, R. A. M. Wolters, and J. G. E. Gardeniers, "The development of titanium silicide–boron-doped polysilicon resistive temperature sensors," *Journal of Micromechanics and Microengineering*, vol. 21, no. 10, pp. 105 022–105 036, Sep. 2011.

[2.9] S. Pan, Y. Luo, S. Heidary Shalmany, and K. A. A. Makinwa, "A resistor-based temperature sensor with a 0.13 pJ·K^2 resolution FoM," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 1, pp. 164–173, 2018.

[2.10] S. Pan and K. A. A. Makinwa, "A 0.25 mm^2 resistor-based temperature sensor with an inaccuracy of 0.12 °C (3σ) from −55 °C to 125 °C," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 12, pp. 3347–3355, 2018.

[2.11] J. Park, J. Kim, K. Kim, J.-H. Yang, M. Choi, and J. Shin, "A 0.65V 1316μm^2 fully synthesizable digital temperature sensor using wire metal achieving 0.16nJ.%^{2}-accuracy FoM in 5nm FinFET CMOS," in *2022 IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 65, 2022, pp. 220–222.

[2.12] M. Shahmohammadi, K. Souri, and K. A. A. Makinwa, "A resistor-based temperature sensor for MEMS frequency references," in *2013 Proceedings of the ESSCIRC (ESSCIRC)*, Sep. 2013, pp. 225–228.

[2.13] P. Park, D. Ruffieux, and K. A. A. Makinwa, "A thermistor-based temperature sensor for a real-time clock with ±2 ppm frequency stability," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 7, pp. 1571–1580, 2015.

[2.14] Y. Lee, W. Choi, T. Kim, S. Song, K. A. A. Makinwa, and Y. Chae, "A 5800-μm^2 resistor-based temperature sensor with a one-point trimmed inaccuracy of ±1.2°C (3σ) from −50°C to 105°C in 65-nm CMOS," *IEEE Solid-State Circuits Letters*, vol. 2, no. 9, pp. 67–70, 2019.

[2.15] A. Khashaba, J. Zhu, A. Elmallah, M. Ahmed, and P. K. Hanumolu, "A 0.0088mm^2 resistor-based temperature sensor achieving 92fJ·K^2 FoM in 65nm CMOS," in *2020 IEEE International Solid-State Circuits Conference - (ISSCC)*, 2020, pp. 60–62.

[2.16] B. Yousefzadeh, U. Sönmez, N. Mehta, J. Borremans, M. A. P. Pertijs, and K. A. A. Makinwa, "A generic read-out circuit for resistive transducers," in *2015 6th International Workshop on Advances in Sensors and Interfaces (IWASI)*, 2015, pp. 122–125.

[2.17] M. H. Perrott, J. C. Salvia, F. S. Lee, A. Partridge, S. Mukherjee, C. Arft, J. Kim, N. Arumugam, P. Gupta, S. Tabatabaei, S. Pamarti, H. Lee, and F. Assaderaghi, "A temperature-to-digital converter for a MEMS-based programmable oscillator with $<$ ±0.5-ppm frequency stability and $<$ 1-ps integrated jitter," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 1, pp. 276–291, 2013.

[2.18] S. Pan, Ç. Gürleyük, M. F. Pimenta, and K. A. A. Makinwa, "A 0.12mm^2 wien-bridge temperature sensor with 0.1°C (3σ) inaccuracy from −40°C to 180°C," in *2019 IEEE International Solid-State Circuits Conference - (ISSCC)*, 2019, pp. 184–186.

[2.19] C. H. Weng, C. K. Wu, and T. H. Lin, "A CMOS thermistor-embedded continuous-time delta-sigma temperature sensor with a resolution FoM of 0.65 pJ $°C^2$," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 11, pp. 2491–2500, Nov. 2015.

[2.20] S. Pan and K. A. A. Makinwa, "A CMOS resistor-based temperature sensor with a 10fJ·K^2 resolution FoM and 0.4°C (3σ) inaccuracy from −55°C to 125°C after a 1-point trim," in *2020 IEEE International Solid-State Circuits Conference - (ISSCC)*, 2020, pp. 68–70.

[2.21] S. Middelhoek, S. Audet, and P. French, *Silicon sensors*. Delft, The Netherlands: Faculty of Information Technology and Systems, Delft University of Technology, 2000.

[2.22] J. Creemer and P. French, "Reduction of uncertainty in the measurement of the piezoresistive coefficients of silicon with a three-element rosette," vol. 3328, 07 1998.

[2.23] F. Fruett and G. C. M. Meijer, *The Piezojunction Effect in Silicon Integrated Circuits and Sensors*. Boston: Kluwer Academic Publishers, May 2002.

[2.24] K. Matsuda, K. Suzuki, K. Yamamura, and Y. Kanda, "Nonlinear piezoresistance effects in silicon," *Journal of Applied Physics*, vol. 73, no. 4, pp. 1838–1847, 1993.

[2.25] Y. Kanda, "A graphical representation of the piezoresistance coefficients in silicon," *IEEE Transactions on Electron Devices*, vol. 29, no. 1, pp. 64–70, Jan. 1982.

[2.26] P. French and A. Evans, "Piezoresistance in polysilicon and its applications to strain gauges," *Solid-State Electronics*, vol. 32, no. 1, pp. 1 – 10, 1989.

[2.27] S. Huber, A. Laville, C. Schott, and O. Paul, "A bridge-type resistive temperature sensor in CMOS technology with low stress sensitivity," in *IEEE SENSORS*, 2014, pp. 1455–1458.

[2.28] U. Ausserlechner, M. Motz, and M. Holliber, "Drift of magnetic sensitivity of smart Hall sensors due to moisture absorbed by the IC-package [automotive applications]," in *IEEE SENSORS*, 2004, pp. 455–458 vol.1.

[2.29] M. Motz, U. Ausserlechner, and M. Holliber, "Compensation of mechanical stress-induced drift of bandgap references with on-chip stress sensor," *IEEE Sensors Journal*, vol. 15, no. 9, pp. 5115–5121, Sep. 2015.

[2.30] D. Manic, J. Petr, and R. S. Popovic, "Short and long-term stability problems of Hall plates in plastic packages," in *IEEE International Reliability Physics Symposium Proceedings*, 2000, pp. 225–230.

[2.31] W. Choi, Y. Lee, S. Kim, S. Lee, J. Jang, J. Chun, K. A. A. Makinwa, and Y. Chae, "A compact resistor-based CMOS temperature sensor with an inaccuracy of 0.12 °C (3σ) and a resolution FoM of 0.43 pJ·K ^{2} in 65-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 12, pp. 3356–3367, 2018.

[2.32] V. Szekely and M. Rencz, "A new monolithic temperature sensor: The thermal feedback oscillator," in *Proceedings of the International Solid-State Sensors and Actuators Conference*, vol. 2, 1995, pp. 124–127.

[2.33] 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," in *2006 IEEE International Solid State Circuits Conference - Digest of Technical Papers*, 2006, pp. 1141–1150.

[2.34] Y. Touloukian and D. DeWitt, *Thermophysical Properties of Matter, Vol. 8, Thermal Radiative Properties: Nonmetallic Solids*, Jan. 1972, vol. 8.

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

[2.36] K. A. A. Makinwa and J. F. Witte, "A temperature sensor based on a thermal oscillator," in *IEEE SENSORS*, Oct. 2005.

[2.37] S. M. Kashmiri, S. Xia, and K. A. A. Makinwa, "A temperature-to-digital converter based on an optimized electrothermal filter," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 7, pp. 2026–2035, July 2009.

[2.38] U. Sönmez, F. Sebastiano, and K. A. A. Makinwa, "Compact thermal-diffusivity-based temperature sensors in 40-nm CMOS for SoC thermal monitoring," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 3, pp. 834–843, 2017.

[2.39] C. P. L. van Vroonhoven, D. d'Aquino, and K. A. A. Makinwa, "A thermal-diffusivity-based temperature sensor with an untrimmed inaccuracy of ±0.2 °C (3σ) from −55 °C to 125 °C," in *2010 IEEE International Solid-State Circuits Conference - (ISSCC)*, 2010, pp. 314–315.

[2.40] C. V. Vroonhoven, *Integrated Temperature Sensors based on Heat Diffusion*, ser. Ph.D. dissertation. Delft, the Netherlands: Delft University of Technology, 2015.

[2.41] S. Pan, J. A. Angevare, and K. A. A. Makinwa, "A hybrid thermal-diffusivity/resistor-based temperature sensor with a self-calibrated inaccuracy of ±0.25°C from −55°C to 125°C," in *2021 IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 64, 2021, pp. 78–80.

[2.42] X. Li, K. Maute, M. Dunn, and R. Yang, "Strain effects on the thermal conductivity of nanostructures," *Physical Review B*, vol. 81, June 2010.

[2.43] M. K. Law, A. Bermak, and H. C. Luong, "A sub-μW embedded CMOS temperature sensor for RFID food monitoring application," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 6, pp. 1246–1255, June 2010.

[2.44] H. Wang and P. P. Mercier, "A 763 pW 230 pJ/Conversion fully integrated CMOS temperature-to-digital converter with +0.81°C/−0.75°C inaccuracy," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 8, pp. 2281–2290, 2019.

[2.45] M. Cochet, B. Keller, S. Clerc, F. Abouzeid, A. Cathelin, J.-L. Autran, P. Roche, and B. Nikoli, "A 225μm^2 probe single-point calibration digital temperature sensor using body-bias adjustment in 28nm FD-SOI CMOS," *IEEE Solid-State Circuits Letters*, vol. 1, no. 1, pp. 14–17, 2018.

[2.46] A. Vaz, A. Ubarretxena, I. Zalbide, D. Pardo, H. Solar, A. Garcia-Alonso, and R. Berenguer, "Full passive UHF tag with a temperature sensor suitable for human body temperature monitoring," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 57, no. 2, pp. 95–99, 2010.

[2.47] A. Annema, "Low-power bandgap references featuring DTMOSTs," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 7, pp. 949–955, 1999.

[2.48] K. Souri, Y. Chae, Y. Ponomarev, and K. A. A. Makinwa, "A precision DTMOST-based temperature sensor," in *2011 Proceedings of the ESSCIRC*, 2011, pp. 279–282.

[2.49] K. Souri, Y. Chae, F. Thus, and K. Makinwa, "A 0.85V 600nW all-CMOS temperature sensor with an inaccuracy of ±0.4°C (3σ) from −40 to 125°C," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2014, pp. 222–223.

[2.50] P. Chen, C. Chen, C. Tsai, and W. Lu, "A time-to-digital-converter-based CMOS smart temperature sensor," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 8, pp. 1642–1648, 2005.

[2.51] W. Zhao, R. Pan, Y. Ha, and Z. Yang, "A 0.4V 280-nW frequency reference-less nearly all-digital hybrid domain temperature sensor," in *2014 IEEE Asian Solid-State Circuits Conference (A-SSCC)*, 2014, pp. 301–304.

[2.52] D. Ha, K. Woo, S. Meninger, T. Xanthopoulos, E. Crain, and D. Ham, "Time-domain CMOS temperature sensors with dual delay-locked loops for microprocessor thermal monitoring," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, no. 9, pp. 1590–1601, 2012.

[2.53] S. Hwang, J. Koo, K. Kim, H. Lee, and C. Kim, "A 0.008 mm^2 500 μW 469 kS/s frequency-to-digital converter based CMOS temperature sensor with process variation compensation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 9, pp. 2241–2248, 2013.

[2.54] T. Someya, A. K. M. M. Islam, T. Sakurai, and M. Takamiya, "An 11-nW CMOS temperature-to-digital converter utilizing sub-threshold current at sub-thermal drain voltage," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 3, pp. 613–622, 2019.

[2.55] T. Someya, A. K. M. M. Islam, and K. Okada, "A 6.4 nW 1.7% relative inaccuracy CMOS temperature sensor utilizing sub-thermal drain voltage stabilization and frequency-locked loop," *IEEE Solid-State Circuits Letters*, vol. 3, pp. 458–461, 2020.

[2.56] K. Yang, Q. Dong, W. Jung, Y. Zhang, M. Choi, D. Blaauw, and D. Sylvester, "A 0.6nJ $-0.22/+0.19$ °C inaccuracy temperature sensor using exponential subthreshold oscillation dependence," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, 2017, pp. 160–161.

[2.57] Z. Tang, Y. Fang, Z. Shi, X.-P. Yu, N. N. Tan, and W. Pan, "A 1770-μm^2 leakage-based digital temperature sensor with supply sensitivity suppression in 55-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 3, pp. 781–793, 2020.

[2.58] Z. Tang, Z. Huang, X.-P. Yu, Z. Shi, and N. N. Tan, "A 0.26-pJ·K^2 2400-μm^2 digital temperature sensor in 55-nm CMOS," *IEEE Solid-State Circuits Letters*, vol. 4, pp. 96–99, 2021.

[2.59] T. Anand, K. A. Makinwa, and P. K. Hanumolu, "A self-referenced VCO-based temperature sensor with 0.034°C/mV supply sensitivity in 65nm CMOS," in *2015 Symposium on VLSI Circuits (VLSI Circuits)*, 2015, pp. C200–C201.

[2.60] K. A. A. Makinwa, *Temperature Sensor Performance Survey,*, [online]. Available: http://ei.ewi.tudelft.nl/docs/TSensor_survey.xls.

[2.61] M. K. Law and A. Bermak, "A 405-nW CMOS temperature sensor based on linear MOS operation," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 56, no. 12, pp. 891–895, 2009.

[2.62] S. Park, Y. Kim, W. Choi, Y. Lee, S. Kim, Y. Shin, and Y. Chae, "A DTMOST-based temperature sensor with 3σ inaccuracy of ± 0.9°C for self-refresh control in 28nm mobile DRAM," in *2020 IEEE Custom Integrated Circuits Conference (CICC)*, 2020, pp. 1–4.

[2.63] D. Esseni, F. Conzatti, M. De Michielis, N. Serra, P. Palestri, and L. Selmi, "Semiclassical transport modelling of CMOS transistors with arbitrary crystal orientations and strain engineering," *Journal of Computational Electronics*, vol. 8, pp. 209–224, 10 2009.

[2.64] N. Wacker, H. Richter, M.-U. Hassan, H. Rempp, and J. N. Burghartz, "Compact modeling of CMOS transistors under variable uniaxial stress," *Solid-State Electronics*, vol. 57, no. 1, pp. 52 – 60, 2011.

[2.65] M. A. P. Pertijs and J. H. Huijsing, *Precision temperature sensors in CMOS technology*. Dordrecht, The Netherlands: Springer, Jan. 2006.

[2.66] E. A. Vittoz, "MOS transistors operated in the lateral bipolar mode and their application in CMOS technology," *IEEE Journal of Solid-State Circuits*, vol. 18, no. 3, pp. 273–279, 1983.

[2.67] M. Eberlein and H. Pretl, "Recent developments in bandgap references for nanometer CMOS technologies," in *2020 Austrochip Workshop on Microelectronics (Austrochip)*, 2020, pp. 42–46.

[2.68] M. Eberlein and H. Pretl, "A no-trim, scaling-friendly thermal sensor in 16nm FinFET using bulk diodes as sensing elements," *IEEE Solid-State Circuits Letters*, vol. 2, no. 9, pp. 63–66, 2019.

[2.69] Z. Tang, Y. Fang, X.-P. Yu, N. N. Tan, Z. Shi, and P. Harpe, "An energy-efficient capacitively biased diode-based temperature sensor in 55-nm CMOS," *IEEE Solid-State Circuits Letters*, vol. 4, pp. 210–213, 2021.

[2.70] Z. Tang, S. Pan, and K. A. A. Makinwa, "A sub-1V 810nW capacitively-biased bjt-based temperature sensor with an inaccuracy of ± 0.15°C (3σ) from -55°C to 125°C," in *2023 IEEE International Solid-State Circuits Conference (ISSCC)*, 2023, pp. 22–24.

[2.71] E. Hellen, "Verifying the diodecapacitor circuit voltage decay," *American Journal of Physics*, vol. 71, p. 797, 2003.

[2.72] M. A. P. Pertijs, K. A. A. Makinwa, and J. H. Huijsing, "A CMOS smart temperature sensor with a 3σ-inaccuracy of ± 0.1°C from -55°C to 125°C," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 12, pp. 2805–2815, Dec. 2005.

[2.73] 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°C to 125°C," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 1, pp. 292–301, Jan. 2013.

[2.74] F. Sebastiano, L. J. Breems, K. A. A. Makinwa, S. Drago, D. M. W. Leenaerts, and B. Nauta, "A 1.2-V 10-μW NPN-based temperature sensor in 65-nm CMOS with an inaccuracy of 0.2°C (3σ) from -70°C to 125°C," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 12, pp. 2591–2601, Dec. 2010.

[2.75] B. Yousefzadeh, S. Heidary Shalmany, and K. A. A. Makinwa, "A BJT-based temperature-to-digital converter with ±60 mk ($3\ \sigma$) inaccuracy from -55°C to $+125$°C in 0.16-μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 4, pp. 1044–1052, April 2017.

[2.76] S. H. Shalmany, K. Souri, U. Sönmez, K. Souri, M. D'Urbino, S. Hussaini, D. Tauro, and S. Tabatabaei, "A 620μW BJT-based temperature-to-digital converter with 0.65mK resolution and FoM of 190fJ·K^2," in *2020 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2020, pp. 70–72.

[2.77] B. Wang, M. Law, A. Bermak, and H. C. Luong, "A passive RFID tag embedded temperature sensor with improved process spreads immunity for a -30°C to 60°C sensing range," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 2, pp. 337–346, 2014.

[2.78] G. Wang, A. Heidari, K. A. A. Makinwa, and G. C. M. Meijer, "An accurate BJT-based CMOS temperature sensor with duty-cycle-modulated output," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 2, pp. 1572–1580, 2017.

[2.79] B. Wang, M. Law, C. Tsui, and A. Bermak, "A 10.6 pJ·K^2 resolution FoM temperature sensor using astable multivibrator," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 7, pp. 869–873, 2018.

[2.80] Z. Tang, Y. Fang, X. Yu, Z. Shi, and N. Tan, "Capacitor-reused CMOS temperature sensor with duty-cycle-modulated output and 0.38°C (3σ) inaccuracy," *Electronics Letters*, vol. 54, no. 9, pp. 568–570, 2018.

[2.81] J. S. Shor and K. Luria, "Miniaturized BJT-based thermal sensor for microprocessors in 32- and 22-nm technologies," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 11, pp. 2860–2867, 2013.

[2.82] K. Souri and K. A. A. Makinwa, *Energy-efficient smart tempearture sensors in CMOS technology*, 1st ed., ser. Analog Circuits and Signal Processing. Springer International Publishing AG, 2018.

[2.83] N. G. Toth, Z. Tang, T. Someya, S. Pan, and K. A. A. Makinwa, "A BJT-based temperature sensor with ±0.1°C(3σ) inaccuracy from -55°C to 125°C and a 0.85pJ·K^2 resolution FoM using continuous-time readout," in *2023 IEEE International Solid-State Circuits Conference (ISSCC)*, 2023, pp. 358–360.

[2.84] C.-Y. Lu, S. Ravikumar, A. D. Sali, M. Eberlein, and H.-J. Lee, "An 8b subthreshold hybrid thermal sensor with ±1.07°C inaccuracy and single-element remote-sensing technique in 22nm FinFET," in *2018 IEEE International Solid-State Circuits Conference - (ISSCC)*, 2018, pp. 318–320.

[2.85] M. Eberlein and I. Yahav, "A 28nm CMOS ultra-compact thermal sensor in current-mode technique," in *2016 IEEE Symposium on VLSI Circuits (VLSI-Circuits)*, 2016, pp. 1–2.

[2.86] A. Aprile, M. Folz, D. Gardino, P. Malcovati, and E. Bonizzoni, "A 0.06-mm^2 current-mode noise-shaping SAR based temperature-to-digital converter with a 4.9-nJ energy/conversion," in *2023 IEEE Custom Integrated Circuits Conference (CICC)*, 2023, pp. 1–2.

[2.87] Y. Li, D. E. Duarte, and Y. Fan, "A 90.9kS/s, 0.7nJ/Conversion hybrid temperature sensor in 4nm-class CMOS," in *2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)*, 2022, pp. 118–119.

[2.88] D. S. Lin and H. P. Hong, "A 0.5V BJT-based CMOS thermal sensor in 10-nm FinFET technology," in *2017 IEEE Asian Solid-State Circuits Conference (A-SSCC)*, 2017, pp. 41–44.

[2.89] J.-H. Park, J.-H. Hwang, C. Shin, and S.-J. Kim, "A BJT-based temperature-to-frequency converter with ±1°c (3σ) inaccuracy from −40°C to 140°C for on-chip thermal monitoring," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 10, pp. 2909–2918, 2022.

[2.90] B. Yousefzadeh and K. A. A. Makinwa, "A BJT-based temperature-to-digital converter with a ±0.25°C 3σ-inaccuracy from −40°C to +180°C using heater-assisted voltage calibration," *IEEE Journal of Solid-State Circuits*, pp. 1–9, 2019.

[2.91] B. Wang, M.-K. Law, and A. Bermak, "A BJT-based CMOS temperature sensor achieving an inaccuracy of ±0.45°C(3σ) from −50°C to 180°C and a resolution-FoM of 7.2pJ·K^2 at 150°C," in *2022 IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 65, 2022, pp. 72–74.

[2.92] U. Kamath, T. Yu, W. Yao, E. Cullen, J. Jennings, S. Wu, P. Lim, B. Farley, and R. B. Staszewski, "BJT device and circuit co-optimization enabling bandgap reference and temperature sensing in 7-nm FinFET," in *2018 48th European Solid-State Device Research Conference (ESSDERC)*, 2018, pp. 86–89.

[2.93] Y. Kim, W. Choi, J. Kim, S. Lee, S. Lee, H. Kim, K. A. A. Makinwa, Y. Chae, and T. W. Kim, "A 0.02mm^2 embedded temperature sensor with ±2 °C inaccuracy for self-refresh control in 25nm mobile DRAM," in *41st European Solid-State Circuits Conference (ESSCIRC)*, 2015, pp. 267–270.

[2.94] G. Sun, Y. Sun, T. Nishida, and S. Thompson, "Hole mobility in silicon inversion layers: Stress and surface orientation," *Journal of Applied Physics*, vol. 102, pp. 084 501–084 501, Oct. 2007.

[2.95] S. Suthram, J. C. Ziegert, T. Nishida, and S. E. Thompson, "Piezoresistance coefficients of (100) silicon nMOSFETs measured at low and high (∼1.5 GPa) channel stress," *IEEE Electron Device Letters*, vol. 28, no. 1, pp. 58–61, 2007.

[2.96] M. Chu, Y. Sun, U. Aghoram, and S. Thompson, "Strain: A solution for higher carrier mobility in nanoscale MOSFETs," *Annual Review of Materials Research*, vol. 39, pp. 203–229, 07 2009.

[2.97] N. Xu, *Effectiveness of Strain Solutions for Next-Generation MOSFETs*. University of California, Berkeley, 2012.

Chapter 3

Bipolar junction transistors in CMOS technology

BJTs are the most suitable candidate for realizing accurate and low-cost TDCs. In a standard CMOS process, BJTs can be realized as parasitic devices, which are mainly used to implement bandgap references and temperature sensors. This chapter focuses on the physics of BJT devices, and in particular how this can be exploited to realize accurate temperature sensors.

3.1 Ideal $I_C - V_{BE}$ characteristic

For a PNP device in the forward active region, the collector current, I_C, is mainly determined by the minority carriers injected into the base from the emitter in the BE junction, I_{pE}, which diffuses through the base to the collector. When the diffusion length L_B is larger than the base width W_B, then I_C will be determined by V_{BE}. As derived in the A.1, I_C can be expressed as:

$$I_C = I_S \cdot \left(exp(\frac{qV_{BE}}{kT}) - 1 \right) , \tag{3.1}$$

where k is the Boltzmann constant, T is the absolute temperature in Kelvin and q is the electron charge in Coulomb. I_S is known as the saturation current, which as derived in A.1, can be expressed as:

$$I_S = A_E q \frac{D_B}{W_B} \frac{n_i^2}{N_B} . \tag{3.2}$$

D_B is the diffusion constant in the base region, A_E is the emitter area, n_i is the intrinsic concentration of the base material (i.e., silicon) and N_B is the base doping.

In deriving 3.1 and 3.2 all other current components are neglected, which means that $I_C = I_E$ is only determined by V_{BE}. In practice, $I_C \neq I_E$ due to non-zero base current, and also I_C depends on V_{CB} (see A.2, A.4).

3.2 Saturation current (I_S)

3.2.1 Temperature dependence of I_S

I_S is a strong function of temperature due to the temperature dependence of the parameters in (3.2): n_i, D_B, and W_B. As will be described A.3, $n_i^2(T)$ is a function of the energy gap between the valence and conduction bands of silicon, $E_g(T)$, which itself changes linearly with temperature. Furthermore, $D_B(T)$ is a function of mobility of the minority carriers in the base. μ is also temperature-dependent (see Fig. 2.1), which is often modelled with a power function: $\mu(T) \propto T^{-m}$. Taking μ and E_g into account and assuming that W_B is fixed, the temperature dependence of I_S can be approximated as:

$$I_S(T) \approx C \cdot T^{\eta} \cdot exp\left(\frac{-qV_{g0}}{kT}\right), \tag{3.3}$$

where V_{g0} ($= E_{g0}/q$) is the bandgap voltage of silicon at 0 K, η is a constant that is related to the temperature-dependence of mobility, and C is a fitting parameter.

3.2.2 Spread of I_S

The spread of I_S will be determined by the parameters in (3.2). N_B varies due to fabrication tolerances, and may spread from -11% to $+13\%$ [3.1]. This is one of the main sources of spread in I_S. n_i^2 is not a process-dependent parameter and can be assumed to be fixed for pure silicon. In a vertical PNP, A_E is determined by the lithographic accuracy, improving quadratically with process scaling. For vertical PNPs, W_B is defined by both lithography and by the depth of the n-well and p+ regions. For lateral PNPs, however, it is determined by lithography and thus improves with process scaling. In lateral BJTs, the parasitic vertical BJT contributes to the collector current, and therefore, needs to be taken into account.

If we assume that the temperature dependence of μ_p and n_i does not spread, the

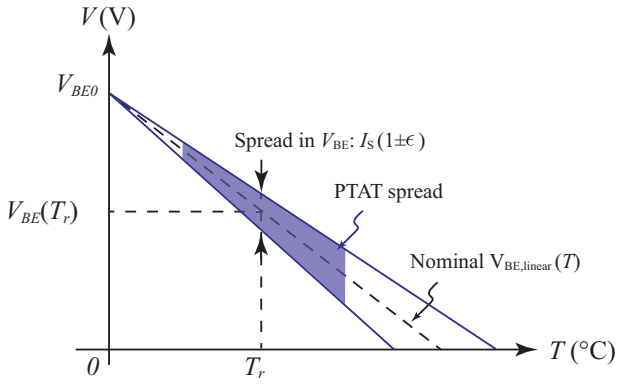


Figure 3.1: PTAT error of V_{BE} as a result of I_S spread. The non-linearity of V_{BE} is not included.

major sources of spread of I_S are the spread of N_B and W_B. These variations only change the absolute value of I_S and not its temperature dependence. As a result, their effect on I_S can be expressed as $I'_S = I_S(1+\epsilon)$, where ϵ models the combined effect of both spread sources. This causes an error in V_{BE} of:

$$V_{BE} = \frac{kT}{q} \cdot ln\left(\frac{I_C}{I_S \cdot (1+\epsilon)}\right) \approx \frac{kT}{q} \cdot ln(\frac{I_C}{I_S}) + \frac{kT}{q} \cdot \epsilon\,, \tag{3.4}$$

which becomes zero at 0 K and increases with temperature. The key observation from (3.4) is that the resulting error in V_{BE} is PTAT, which can be corrected with a single point trim.

3.2.3 I_S compensation

Methods to correct the spread of I_S without using any external calibration references have been suggested in previous work. These methods typically use a parameter that is highly correlated with I_S to determine and then eliminate the spread in I_S [3.2]–[3.5]. A high correlation between n-well resistance and I_S has been observed since both are affected by the variation in N_B [3.1]. Using a pinched base resistance to compensate for the spread of I_S has been suggested in [3.2],[3.3]. Using another BJT in saturation, i.e., both V_{BE} and V_{BC} forward-biased, has also been suggested. In this method, the recombination current in the base of a BJT that is in saturation is used to correct the spread of the I_S in the forward-active region [3.4].

3.3 Current gains α_F and β

Not all carriers injected into the Base from the Emitter can reach the Collector. Considering other current components results in $I_E \neq I_C$ (see A.2). To account for this, $\alpha_F = I_C/I_E$ is defined, which can be expressed as the multiplication of three terms: the *emitter injection efficiency* γ, the *base transport factor* α_T, and the *recombination factor* δ. For circuit design analysis, the current gain $\beta = I_C/I_B$ is often used, which is derived from $\beta = \alpha_F/(1-\alpha_F)$.

When W_B is smaller than L_B, $\alpha_T \approx 1$. Assuming $\delta \approx 1$, β is mainly determined by the emitter injection efficiency and can be expressed as:

$$\beta = \frac{D_B W_E N_E}{D_E W_B N_B}\,. \tag{3.5}$$

The key observation is that β is independent of electrical conditions and is determined by the relative dimension and doping concentration between the base and the emitter. In high-performance bipolar devices, W_B is made very small, resulting in a high α_F of ~ 0.998. For vertical PNPs in CMOS technology, this is not the case due to the large W_B, which may result in α_F of 0.5 or even lower [3.6].

When V_{BE} is very small, due to recombination-generation in the BE junction, the assumption of $\delta \approx 1$ is not valid. Due to the lower δ, α_F drops, thereby causing β to drop. For very high values of V_{BE}, β drops again, but due to a different mechanism known as the *high-injection* effect in the BE junction. In the region between these extreme V_{BE} values, β remains flat, and has values expressed by 3.5. This is the region of interest for use in bandgap voltage references and temperature sensors. The flat region depends on the device geometry and fabrication process. As shown in Fig. 3.2, in 0.16-μm technology, the β versus I_C exhibits a fairly large flat region [3.7], while in 65-nm technology, this region is quite small [3.8].

3.3.1 Temperature dependence of β

The temperature-dependence of β is empirically modeleld as $\beta(T) = \beta(T_r)(T/T_r)^{XTB}$, where T_r is an arbitrary reference temperature, and XTB is a process-dependent constant ranging from 1.5 to 2.5.

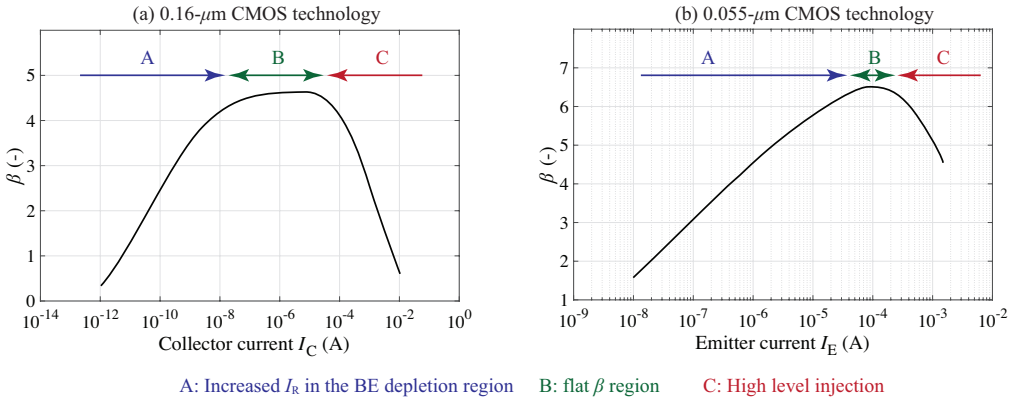


Figure 3.2: Simulated β of vertical PNPs in: (a) 0.16-μm CMOS technology taken from [3.7] and (b) 0.065-μm technology taken from [3.8].

3.3.2 Spread of β

In the flat region, the variation of β is related to the spread of the parameters in (3.5): W_E, W_B, N_B, N_E, as well as μ in the base and in the emitter regions. As will be shown in Chapter 4, the spread of β can be $\pm 10\%$. Since vertical PNPs must be biased via their emitter terminal, their collector current I_C $(= I_E \cdot \beta/(\beta+1))$ and the resulting V_{BE} will be β-dependent, which can be expressed as:

$$V_{BE} = \frac{kT}{q} ln(\frac{I_E}{I_S}) + \frac{kT}{q} ln(\frac{\beta}{\beta+1}) . \tag{3.6}$$

The spread of β causes V_{BE} to vary. Assuming XTB is fixed, and considering the spread of β_0, V_{BE} can be expressed as:

$$V_{BE}(\beta_0 + \Delta\beta_0) \approx V_{BE}(\beta_0) - \frac{kT}{q} \cdot \frac{\Delta\beta_0^2}{\beta_0^2} \cdot \left(1 + \frac{1}{\beta_0} \cdot (\frac{T}{T_r})^{-XTB}\right)^{-1} . \tag{3.7}$$

The key observation from (3.7), is that when $\beta_0 \gg 1$ the resulting error in V_{BE} is PTAT and can thus be trimmed with that of I_S. When β is small, the resulting error is not PTAT, and there will be residual error after the trim. Additionally, the spread of XTB will contribute to the residual error in V_{BE}. Since vertical BJTs have small β values (e.g., $\sim$5), the effect of β on V_{BE} needs to be compensated for with the help of circuit techniques.

The spread of β may also impact the ΔV_{BE} that is derived from two vertical PNPs biased at a fixed emitter current ratio, $I_E : pI_E$. When both of these currents

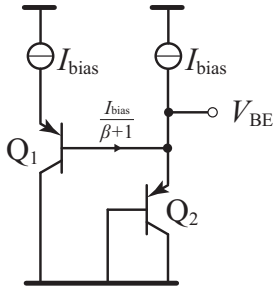


Figure 3.3: β compensation method using the base current of an auxilary PNP [3.9].

are within an entirely flat β region, the effect of β on ΔV_{BE} cancels out. However, differences in β at the two currents introduce an error in ΔV_{BE} that can be expressed as:

$$\Delta V_{BE} = V_{BE}(pI, \beta) - V_{BE}(I, \beta + \Delta\beta) \approx \frac{kT}{q} \ln(p) + \frac{kT}{q} \cdot \left(\frac{\Delta\beta}{\beta^2} \right) . \tag{3.8}$$

The accuracy of ΔV_{BE} is thus a function of β, which has shown a decreasing trend as technology feature sizes shrink. At room temperature, the nominal β for vertical PNPs ranges from 25 in a 0.7-μm [3.1] to 0.9 in a 0.055-μm [3.6] CMOS technology.

3.3.3 β compensation

In previous work, methods for compensating for the effect of β in V_{BE} have been suggested. In [3.9], the base current of an identical PNP is injected into the emitter causing the collector current of the target PNP, Q_2, to be less dependent on β. With reference to Fig. 3.3, this collector current can be written as:

$$I_C = \frac{\beta}{1+\beta} \left(I_{bias} + \frac{I_{bias}}{1+\beta} \right) = I_{bias} \left(\frac{\beta(\beta+2)}{\beta(\beta+2)+1} \right) . \tag{3.9}$$

This β-compensation method essentially increases the effective β to $\beta \cdot (\beta + 2)$ thus reducing the related error in V_{BE}. The main limitation of this method is the minimum supply voltage required, which needs to be well above $2V_{BE}$ for proper operation of current sources and the BJTs. This rules out this method in low-voltage applications.

Another method for compensating for β is to bias the PNP with a β-dependent emitter current $I_E = I_{bias} \cdot \beta/(1+\beta)$. This would then cause I_C, and thus V_{BE} to be β-independent [3.10]. This method will be further discussed in Section 4.3.

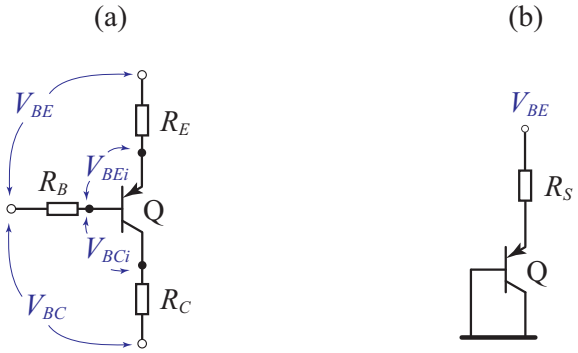


Figure 3.4: (a) Resistances associated with PNP regions. (b) Equivalent series resistance R_S.

3.4 Series resistances (R_S)

Each terminal of a BJT is associated with an equivalent resistance that originates from the resistance of the doped material and the metal interconnects. As shown in Fig. 3.4 (a), the voltage drop across these resistances cause V_{BEi} and V_{BCi} to differ from their externally measured voltages V_{BE} and V_{BC}. R_E is usually small due to high doping level of emitter, while R_B is relatively large due to low doping level. In BiCMOS technologies, R_C is made low, however, in vertical PNPs in CMOS, the collector region can be assumed to have the same doping as the base. The presence of R_C only causes second-order effects via the Early effect by changing V_{BC}. Neglecting R_C, the V_{BE} of a diode-connected PNP can be written as:

$$V_{BE} = V_{BEi} + R_E \cdot I_E + R_B \cdot \frac{I_E}{\beta + 1} = V_{BEi} + R_S \cdot I_E \,, \tag{3.10}$$

where $R_S = R_E + R_B/(1 + \beta)$ is the equivalent series resistance that is modeled in series with the emitter as illustrated in Fig. 3.4 (b).

R_S also affects ΔV_{BE}. For two PNPs with emitter current ratio $I_E : p \cdot I_E$, the resulting error in ΔV_{BE} can be expressed as:

$$\Delta V_{BE} = \frac{kT}{q} ln(p) + (p - 1) \cdot I_E \cdot R_S \,, \tag{3.11}$$

in which β is assumed to be current-independent. It is worth noting that ΔV_{BE} only increases with a logarithmic function when the ratio p increases, while the associated

R_S-related error increases linearly and thus increase faster. This discourages the choice of extremely large current ratios (e.g., 20) for realizing a large ΔV_{BE}.

The effect of R_S can be compensated for by using circuit techniques. Since V_{BE} is logarithmically related to I_E while the voltage drop on R_S is linearly related to I_E, a linear combination of V_{BE} at different currents can remove the effect of R_S [3.1]. This is further discussed in Sec. 4.2.3.

3.4.1 Base resistance (R_B)

Since the base region is made of low-doped material, the flow of current in the base causes a voltage drop in the base region, which then affects the path of carriers that flow across the base. As a result of this effect, known as *current crowding*, the base resistance is modelled as a fixed extrinsic resistance, R_{Bext}, in series with a current-dependent intrinsic resistance, R_{Bint}, as [3.11]:

$$R_B(I_B) = R_{B,ext} + R_{B,int}(I_B) = R_{B,ext} + R_{B,int0} \cdot \frac{I_{B0}}{I_{B0} + I_B}\,, \tag{3.12}$$

where I_{B0} and R_{Bint0} are fitting parameters. In a 0.8-μm BiCMOS technology I_{B0}, R_{Bint0}, and R_{Bext} were measured to be 110 μA, 1.1 kΩ, and 33 Ω, respectively [3.11]. It is worth noting that R_B decreases as I_B increases. In a 0.16-μm CMOS technology, a 5×5 μm^2 PNP device exhibited an R_B of 60 Ω at an emitter current of 1 mA which then increased to 600 Ω at an emitter current of 1 μA [3.12].

The temperature-dependence of R_B is assumed to be similar to that of n-well resistors. In 0.7- and 0.5-μm technologies, an R_B of $\sim$ 500 Ω and $\sim$ 125 Ω was measured at room temperature, which then increased to $\sim$ 1000 Ω and $\sim$ 200 Ω at 100 °C [3.13].

3.5 Effect of mechanical stress on BJTs

Mechanical stress can affect the performance of semiconductor devices. Stress can originate from thermo-mechanical effects during silicon wafer processing, or due to packaging. This occurs because the temperature during the wafer fabrication processes (up to $\sim$1200°C) and packaging (up to $\sim$200°C) is higher than the operating temperature of the circuit. When the fabricated silicon die cools down, the differences in the thermal expansion coefficient (TEC) of materials give rise to mechanical stress. Stress resulting from the fabrication process originate from different mecha-

nisms, which can be further studied in [3.14]. These are often unavoidable and can only be reduced, for instance, by keeping critical devices away from the edges of the silicon die, using stress-insensitive layout techniques, or by avoiding the placement of stress-generating films near critical devices.

3.5.1 Stress due to packaging

Packaging is necessary to provides mechanical support and also resistance to humidity, ionic contamination, and radiation for electronic devices. After fabrication, the wafer is diced, and the silicon die will attach to a die-paddle. It is then wire bonded to lead frame pads, passivated, and finally either encapsulated in plastic packaging or hermetically sealed in ceramic packaging. The use of plastic packaging accounts for more than 99% of device packagings. The use of ceramic packaging, due to their higher costs ($\sim 10\times$), is limited to low-stress applications and MEMS components [3.15].

In ceramic packaging, silicon die is attached to a die-paddle using an adhesive, and then a metal lid is placed on top and sealed by welding or using adhesives (see Fig. 3.5 (a)). The resulting stress is due to the difference between the TEC of the silicon die ($\sim\ 2.5 - 4.5 \times 10^{-6} K^{-1}$), the die attachment material that is either an epoxy adhesive or polyamide (TEC $\sim\ 4060 \times 10^{-6} K^{-1}$) and the lead frame that is made of copper, or an alloy of Ni-Fe (TEC $\sim\ 4 - 17 \times 10^{-6} K^{-1}$). This creates *tensile* in-plane stress of about 35-55 MPa [3.1] on the surface of the die. In plastic encapsulation, the die and the die paddle is placed in a cavity, which is then filled with a thermosetting polymer. The polymer become solid at its curing temperature and remains solid. This causes an additional *compressive* stress in the order of 200 MPa [3.1],[3.16] due to the TEC difference of the die and the encapsulant material that is usually an epoxy resin (TEC $\sim\ 1320 \times 10^{-6} K^{-1}$). In recent packaging technologies, solder bumps are added on top of the die surface during the last step of the fabrication. Die is then flipped and bumps are used to make connection to lead-frame pads. These solder bumps can add additional local stress. Further reading about the materials used in various packaging methods and their TEC can be found in [3.15].

One major concern about plastic packaging is their specific features of the encapsulant materials, which subject them to hysteresis, relaxation, or creep [3.16]. While silicon crystal does not show any of these characteristics, the epoxy material might undergo a time-dependent variation, which then results in different stress applied to the silicon die. Such variation can also be due to swelling of the epoxy material as a result of moisture absorption [3.17].

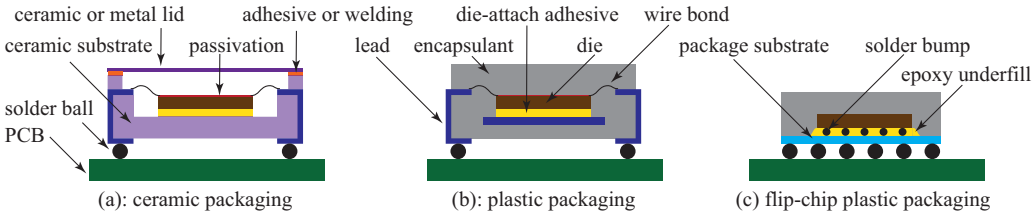


Figure 3.5: Simplified illustration of microelectronic packaging. (a) Ceramic packaging, (b) plastic encapsulation and (c) plastic encapsulation with a flip-chip bonding.

The effect of mechanical stress on temperature sensor can be studied based on its effect on resistors and BJTs. The former is due to the piezo-resistive effect and the latter is due to the piezo-junction effect.

3.5.2 Piezo-junction effect

The variation of the saturation current of a BJT due to applied stress is known as piezo-junction effect. Similar to piezo-resistive effect, the relationship between I_S variation $\delta I_S/I_{S0}$ and applied stress σ is found by determining a set of piezo-junction coefficients. These coefficients are also empirically found since they are dependent on several parameters such as the doping concentration, doping type, and the direction of current flow in the device [3.16].

As a result of stress, value of V_{BE} changes, which can be expressed as:

$$V_{BE}(\sigma) \approx V_{BE}(\sigma = 0) - \frac{kT}{q}\frac{\delta I_S}{I_{S0}}, \tag{3.13}$$

where δI_S is the amount of I_S change due to stress, and I_{S0} is the saturation current under a stress-free condition. Unlike the spread-related errors in V_{BE}, the stress-related errors caused by packaging are not PTAT. This is because the amount of stress σ, and thus $\delta I_S/I_{S0}$ changes over temperature. Since the curing temperature of die attachment and the encapsulant material are high, stress increases at lower temperatures and causes more error on V_{BE}. Therefore, the effect stress cannot be corrected with a PTAT trim.

Fig. 3.6 (a) shows the variation in V_{BE} for both compressive and tensile stress from -180 MPa to $+180$ MPa in both vertical NPN and PNP devices [3.16]. In a PNP device, 200 MPa of compressive stress can cause an error of 1.8 mV on V_{BE} at -10

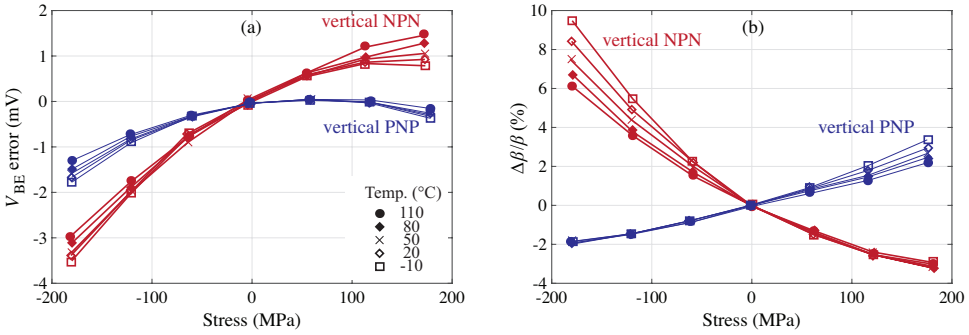


Figure 3.6: Stress-induced variation in (a) V_{BE} and (b) β for vertical BJTs, taken from [3.16]. The PNP had an $A_E = 40 \times 40\ \mu m^2$ and the NPN had an $A_E = 16 \times 2\ \mu m^2$, and both were realized in a 0.8 μm BiCMOS technology.

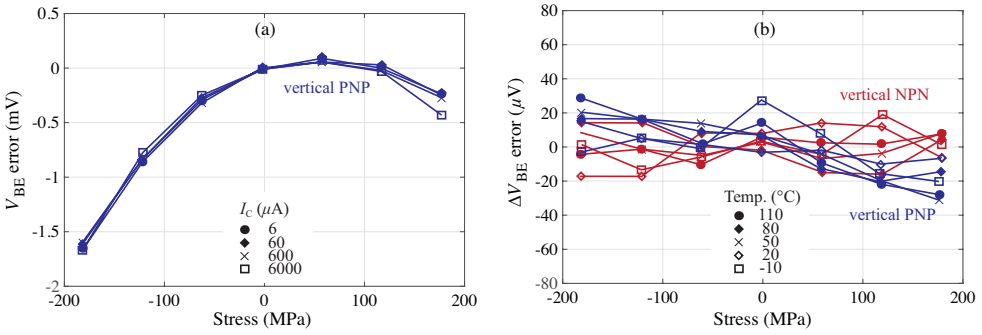


Figure 3.7: (a) Stress-induced effect on V_{BE} at different current levels. (b) Stress-induced effect on ΔV_{BE} for both vertical PNPs and NPNs, taken from [3.16].

°C, while 45 MPa of tensile stress only results in an error of 0.1 mV. A key observation from these measured results is that both types of devices are ($\sim$ 3×) more sensitive to compressive stress than tensile stress. Another important observation is that NPN devices are more stress-sensitive than PNP devices. Besides these observations, it has also been shown in [3.16],[3.18] that vertical BJTs are less stress sensitive than lateral BJTs. Due to the combined effect of the lower amount of stress and the lower stress sensitivity, the effect of stress in ceramic packaging can be assumed to be $\sim$ 15× lower than that of plastic encapsulation.

In addition to the I_S variation, stress can also change the β of BJTs [3.1],[3.16]. However, since the effect of β in V_{BE} can be compensated for by circuit techniques, its stress-related errors will also be canceled. Fig. 3.6 (b) shows the effect of stress on β for both vertical NPN and PNP devices. It is interesting to note that the sensitivity

of β to compressive and tensile stress has opposite signs for NPN and PNP devices.

ΔV_{BE} is less stress sensitive than V_{BE} because the effect of stress on two BJTs at different current densities is nearly equal. As illustrated in Fig. 3.7 (a), the piezo-junction coefficients are independent of current level [3.16]. Applied stress can also be assumed identical for two devices that are on the same die placed close together. The effect of stress on ΔV_{BE} will thus cancel out. Fig. 3.7 (b) shows this for both vertical NPN and PNP devices [3.16].

3.6 Conclusion

This chapter elaborated on the basic physics and operation of BJTs. Since the accuracy and calibration requirements of BJT-based temperature sensors ultimately rely on the characteristics of BJTs, understanding their basic operation is necessary. The main current components that are involved in the forward-active region of a BJT were discussed. The temperature-dependence, and spread, of V_{BE} were derived based on physical parameters. The effect of current-gain and series resistance on the accuracy of V_{BE} were also discussed. Finally, the effect of mechanical stress is discussed, as this is the ultimate limit on the accuracy of a packaged temperature sensor. Stress-generating mechanisms caused by both the IC fabrication process and packaging were discussed, and the effect of the resulting stress on the characteristics of V_{BE} and ΔV_{BE} are shown based on the available literature.

Bibliography

[3.1] M. A. P. Pertijs and J. H. Huijsing, *Precision temperature sensors in CMOS technology*. Dordrecht, The Netherlands: Springer, Jan. 2006.

[3.2] R. P. Fisk and S. M. R. Hasan, "A calibration-free low-cost process-compensated temperature sensor in 130 nm CMOS," *IEEE Sensors Journal*, vol. 11, no. 12, pp. 3316–3329, Dec. 2011.

[3.3] R. Amador, A. Polanco, H. Hernndez, E. Gonzlez, and A. Nagy, "Technological compensation circuit for accurate temperature sensor," *Sensors and Actuators A: Physical*, vol. 69, no. 2, pp. 172 – 177, Aug. 1998.

[3.4] B. Wang, M. K. Law, A. Bermak, and F. Tang, "BJT process spread compensation utilizing base recombination current in standard CMOS," *IEEE Electron Device Letters*, vol. 36, no. 11, pp. 1111–1113, Nov. 2015.

[3.5] R. Amador, A. Polanco, H. Hernandez, E. Gonzalez, and A. Nagy, "Reducing V_{BE} wafer spread of bipolar transistor via a compensation circuit," *Electronics Letters*, vol. 28, no. 15, pp. 1378–1379, July 1992.

[3.6] Z. Tang, Y. Fang, Z. Huang, X. Yu, Z. Shi, and N. N. Tan, "An untrimmed BJT-based temperature sensor with dynamic current-gain compensation in 55-nm CMOS process," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 10, pp. 1613–1617, Oct 2019.

[3.7] K. Souri and K. A. A. Makinwa, "A 0.12 mm^2 7.4 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.

[3.8] X. Pu, M. Ash, K. Nagaraj, J. Park, S. Vu, P. Kimelman, and S. de la Haye, "An embedded 65 nm CMOS remote temperature sensor with digital beta correction and series resistance cancellation achieving an inaccuracy of 0.4 °C (3σ) from -40 °C to 130 °C," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 9, pp. 2127–2137, Sep. 2015.

[3.9] A. Heidary, G. Wang, K. Makinwa, and G. Meijer, "A BJT-based CMOS Temperature Sensor with a 3.6pJK^2-Resolution FoM," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, Feb. 2014, pp. 224–225.

[3.10] M. Pertijs and J. Huijsing, "Bias circuits," Patent, Nov. 4, 2008, uS Patent 7,446,598 B2.

[3.11] M. Linder, F. Ingvarson, K. O. Jeppson, J. V. Grahn, Shi-Li Zhang, and M. Ostling, "Extraction of emitter and base series resistances of bipolar transistors from a single DC measurement," *IEEE Transactions on Semiconductor Manufacturing*, vol. 13, no. 2, pp. 119–126, May 2000.

[3.12] L. Song, H. Homulle, E. Charbon, and F. Sebastiano, "Characterization of bipolar transistors for cryogenic temperature sensors in standard CMOS," in *2016 IEEE SENSORS*, Oct. 2016, pp. 1–3.

[3.13] G. Wang and G. C. Meijer, "The temperature characteristics of bipolar transistors fabricated in CMOS technology," *Sensors and Actuators A: Physical*, vol. 87, no. 1, pp. 81 – 89, Dec. 2000.

[3.14] S. Hu, "Stress related problems in silicon technology," *Journal of Applied Physics*, vol. 70, pp. R53–R80, Oct. 1991.

[3.15] H. Ardebili and M. G. Pecht, *Encapsulation Technologies for Electronic Applications*, 2nd ed., ser. Materials and Processes for Electronic Applications. Elsevier, March 2009.

[3.16] F. Fruett and G. C. M. Meijer, *The Piezojunction Effect in Silicon Integrated Circuits and Sensors*. Boston: Kluwer Academic Publishers, May 2002.

[3.17] M. Motz, U. Ausserlechner, and M. Holliber, "Compensation of mechanical stress-induced drift of bandgap references with on-chip stress sensor," *IEEE Sensors Journal*, vol. 15, no. 9, pp. 5115–5121, Sep. 2015.

[3.18] G. C. M. Meijer, *Integrated circuits and components for bandgap references and temperature transducers*. Delft, The Netherlands: Ph.D. dissertation, Delft University of Technology, Mar. 1982.

[3.19] C. Hu, *Modern Semiconductor Devices for Integrated Circuits*. Prentice Hall, 2010.

[3.20] Y. Varshni, "Temperature dependence of the energy gap in semiconductors," *Physica*, vol. 34, pp. 149–154, Dec. 1967.

[3.21] S. Sze and K. Ng, *Physics of Semiconductor Devices*. Wiley, 2006.

[3.22] I. M. Filanovsky and A. Allam, "Mutual compensation of mobility and threshold voltage temperature effects with applications in CMOS circuits," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 7, pp. 876–884, July 2001.

[3.23] I. E. Getreu, *Modeling the Bipolar Transistor*. Elsevier, 1978.

[3.24] T. H. Ning and D. D. Tang, "Method for determining the emitter and base series resistances of bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 31, no. 4, pp. 409–412, Apr. 1984.

[3.25] A. Ohte and M. Yamagata, "A precision silicon transistor thermometer," *IEEE Transactions on Instrumentation and Measurement*, vol. 26, no. 4, pp. 335–341, Dec. 1977.

[3.26] K. E. Kuijk, "A precision reference voltage source," *IEEE Journal of Solid-State Circuits*, vol. 8, no. 3, pp. 222–226, June 1973.

[3.27] Y. P. Tsividis, "Accurate analysis of temperature effects in $I_C - V_{BE}$ characteristics with application to bandgap reference sources," *IEEE Journal of Solid-State Circuits*, vol. 15, no. 6, pp. 1076–1084, Dec. 1980.

[3.28] Y. Varshni, "Temperature dependence of the energy gap in semiconductors," *Physica*, vol. 34, no. 1, pp. 149 – 154, 1967.

[3.29] G. C. M. Meijer and K. Vingerling, "Measurement of the temperature dependence of the $I_C - (V_{BE})$ characteristics of integrated bipolar transistors," *IEEE Journal of Solid-State Circuits*, vol. 15, no. 2, pp. 237–240, Apr. 1980.

[3.30] S. Heidary Shalmany, D. Draxelmayr, and K. A. A. Makinwa, "A ±5 A integrated current-sensing system with ±0.3% gain error and 16 μA offset from −55 °C to +85 °C," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 4, pp. 800–808, Apr. 2016.

[3.31] S. Heidary Shalmany, D. Draxelmayr, and K. A. A. Makinwa, "A ±36 A integrated current-sensing system with a 0.3% gain error and a 400 μA offset from −55 °C to +85 °C," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 4, pp. 1034–1043, Apr. 2017.

[3.32] P. Malcovati, F. Maloberti, C. Fiocchi, and M. Pruzzi, "Curvature-compensated BiCMOS bandgap with 1-V supply voltage," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 7, pp. 1076–1081, July 2001.

[3.33] J. Hebb and K. Jensen, "The effect of patterns on thermal stress during rapid thermal processing of silicon wafers," *IEEE Transactions on Semiconductor Manufacturing,*, vol. 11, pp. 99–107, March 1998.

[3.34] P. J. Cousins and J. E. Cotter, "Misfit dislocations generated during non-ideal boron and phosphorus diffusion and their effect on high-efficiency silicon solar cells," in *IEEE Photovoltaic Specialists Conference*, Jan. 2005, pp. 1047–1050.

[3.35] G. C. M. Meijer, Guijie 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.

[3.36] B. Abesingha, G. A. Rincon-Mora, and D. Briggs, "Voltage shift in plastic-packaged bandgap references," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 49, no. 10, pp. 681–685, Oct. 2002.

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

[3.38] P. French and A. Evans, "Polysilicon strain sensors using shear piezoresistance," *Sensors and Actuators*, vol. 15, no. 3, pp. 257 – 272, 1988.

[3.39] P. French and A. Evans, "Piezoresistance in polysilicon and its applications to strain gauges," *Solid-State Electronics*, vol. 32, no. 1, pp. 1 – 10, 1989.

[3.40] P. Lu, "Low-frequency noise in self-aligned bipolar transistors," *Journal of Applied Physics*, vol. 62, no. 4, pp. 1335–1339, Aug. 1987.

[3.41] T. G. M. Kleinpenning, "Location of low-frequency noise sources in submicrometer bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 39, no. 6, pp. 1501–1506, June 1992.

[3.42] J. C. Costa, D. Ngo, R. Jackson, N. Camilleri, and J. Jaffee, "Extracting $1/f$ noise coefficients for BJT's," *IEEE Transactions on Electron Devices*, vol. 41, no. 11, pp. 1992–1999, Nov. 1994.

[3.43] D. A. Neamen, *Semiconductor physics and devices: basic principles*, 3rd ed. New York: McGraw-Hill, 2003.

[3.44] S. Middelhoek, S. Audet, and P. French, *Silicon sensors*. Delft, The Netherlands: Faculty of Information Technology and Systems, Delft University of Technology, 2000.

[3.45] J. Creemer and P. French, "Reduction of uncertainty in the measurement of the piezoresistive coefficients of silicon with a three-element rosette," vol. 3328, 07 1998.

[3.46] K. Matsuda, K. Suzuki, K. Yamamura, and Y. Kanda, "Nonlinear piezoresistance effects in silicon," *Journal of Applied Physics*, vol. 73, no. 4, pp. 1838–1847, 1993.

Chapter 4

An accurate and process-insensitive BJT-based TDC[1]

In Chapter 1, the need for a low-cost precision TDC in CMOS technologies was explained. Chapter 2 then compared different methods for realizing TDCs and the differences in terms of their calibration and trimming. Although BJT-based TDCs are low-cost and relatively accurate, they are increasingly difficult to realize in the latest CMOS technology nodes, mainly due to the degradation of BJT performance. These difficulties were briefly discussed in Sec. 3.3.

BJT-based TDCs have achieved 3σ-inaccuracies ranging from ± 0.1°C to ± 0.25°C over the military temperature range (-55°C to $+125$°C), when implemented in CMOS technology nodes ranging from 0.7-μm to 0.16-μm [4.1]–[4.5]. This level of accuracy is achieved by combining a set of average calibration parameters, obtained by a batch calibration, with individual calibration parameters, obtained by a room temperature (RT) calibration.

In a 0.7-μm technology, these average parameters have been shown to remain constant over two production batches [4.5]. In a 0.16-μm technology, however, this is not the case [4.2]. Using the average parameters from one batch to trim another batch led to extra errors of about 1°C (see [4.2], Table I). So, for maximum accuracy, these parameters must be determined for every new batch of TDCs. This significantly complicates their production, because determining these parameters requires a multi-temperature batch calibration, which is a time-consuming, and thus expensive process. In addition, the resulting parameters must then be associated with the samples of the

[1] B. Yousefzadeh, S. Heidary Shalmany and K. A. A. Makinwa, "A BJT-based temperature-to-digital converter with ±60 mK (3 σ) inaccuracy from -55 °C to $+125$ °C in 0.16-μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 4, pp. 1044-1052, Apr. 2017.

correct batch, which is a logistical challenge in itself.

This chapter describes the realization of a TDC that achieves an inaccuracy of ± 0.1 °C (3σ) over the military range in a 0.16-μm technology, and over three batches, while using the same average parameters. This robustness to process spread is mainly due to the use of a precision bias circuit, which ensures that the collector current of the sensing PNPs is reproducible. The major remaining sources of error are then due to the effect of process spread on the PNP's saturation current I_S and on the bias resistor R_b, which can both be corrected by a RT trim. Since substrate PNPs must be biased via their emitter, their collector current is also a function of their current gain β, which also suffers from process spread. To investigate this effect, the TDC was also designed to measure β. One hundred TDC samples from three different batches were then measured to validate the effect of batch-to-batch spread. Furthermore, to circumvent the need for a time-consuming RT calibration and trim, alternative techniques such as voltage calibration [4.6],[4.7], and batch trimming were also investigated.

4.1 Operating principle

The heart of the TDC is a bipolar core consisting of two PNPs that are biased at a fixed current density ratio. It produces two temperature-dependent signals, V_{BE} and ΔV_{BE}. V_{BE} has a CTAT characteristic, while ΔV_{BE} ($= V_{BE2} - V_{BE1}$) has a PTAT characteristic. These voltages are applied to an ADC, which digitizes the ratio X_T ($= V_{BE}/\Delta V_{BE}$). While X_T is a non-linear function of temperature [4.2], it can be converted to a linear function $\mu_T = \alpha/(\alpha + X_T)$, where ($\alpha \sim 16$) is an average calibration parameter. The temperature in degrees Celsius, D_{out}, can then be expressed as $D_{out} = A \cdot \mu_T - B$, where A (~ 615) and B (~ 283) are also average calibration parameters.

4.2 Accuracy limits of the bipolar core

The voltages V_{BE} and ΔV_{BE} deviate from the ideal equations in (2.3) and (2.4) due to the finite β, and non-zero series resistance R_S of the substrate BJTs. β-related errors are shown in the the second term in (3.6) and (3.8), while R_S-related errors are shown in the second term in (3.10) and (3.11).

Errors in V_{BE} and ΔV_{BE} caused by R_S and β can be compensated for by implementing a more complex bipolar core [4.8],[4.9]. To investigate if this is necessary, the accuracy of the bipolar core in a 0.16-μm technology was evaluated by simula-

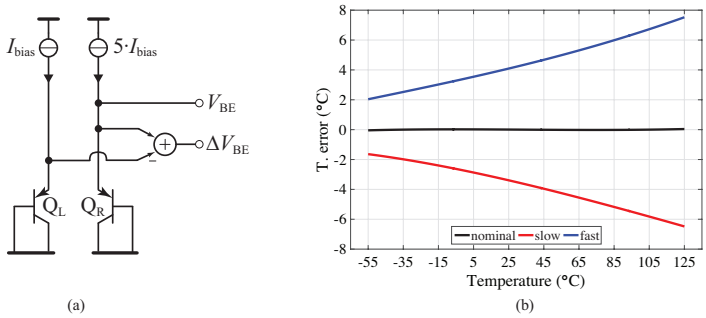


Figure 4.1: (a) Bipolar core without β and R_S compensation. (b) Untrimmed temperature error.

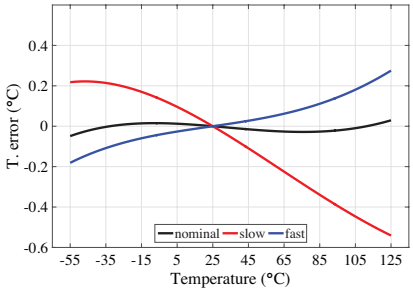


Figure 4.2: Trimmed temperature error for the bipolar core of Fig. 4.1.

tion. The results are discussed in the following section for the following three cases: uncompensated β and R_S, compensated β but uncompensated R_S, and compensated β and R_S.

4.2.1 Uncompensated β and R_S

Fig. 4.1 (a) illustrates a simple bipolar core in which two PNPs (Q_R, Q_L) are biased at a 1 : 5 emitter current ratio, where the unit emitter current is set to I_{bias}. To exclude bias circuit non-idealities, I_{bias} is an ideal PTAT source: $I_{bias} = (kT/q) \cdot ln(5)/(500$ kΩ), which is ~ 80 nA at 25 °C. The PNPs ($5{\times}5$ μm^2) have a nominal $\beta = 4.5$, and $I_S = 9.2 \times 10^{-19}$ at 25 °C. V_{BE} ($= V_{BE2}$) and ΔV_{BE} ($= V_{BE2} - V_{BE1}$) were then simulated from −55 °C to 125 °C at different process corners, and then post-processed to obtain D_{out} and the temperature error $D_{out} - T$. The calibration parameters were obtained at the nominal process corner and then kept constant for the other process corners—a criterion that is also required for low-cost calibration.

As illustrated in Fig. 4.1 (b), the untrimmed temperature error of such a bipolar

core is quite large: about 8 °C. This is mainly due to the large spread of I_S. When only its absolute value is assumed to vary, and not its temperature dependence, this causes a PTAT error in V_{BE} (see (3.30)), or an offset error in X_T as shown in the following:

$$X_T = \frac{V_{BE}}{\Delta V_{BE}} \approx \frac{V_{BE,nom} + \frac{kT}{q} \cdot \epsilon}{\Delta V_{BE}} \approx X_{T,nom} + \frac{\epsilon}{ln(5)}\,, \tag{4.1}$$

in which ϵ is a constant representing error of I_S. The large spread in I_S, can thus be corrected by using a PTAT trim in V_{BE}, or simply by making an offset trim in X_T. Fig. 4.2 shows that the temperature error then drops to ±0.6 °C.

4.2.2 Compensated β and uncompensated R_S

The accuracy of the bipolar core can be enhanced by compensating for the effect of β on V_{BE}. As shown in Fig. 4.3 (a), this can be realized by using ideal circuit elements to copy the base current of Q_R and then add it to the emitter currents of Q_R and Q_L. This is equivalent to the β-compensation methods used in [4.1],[4.2],[4.10],[4.11]. The trimmed spread then improves by 10× to ±0.06 °C, as shown in Fig. 4.4.

This β-compensation method would be even more accurate if the β of the two PNPs were exactly the same. In practice, however, β is also a function of the bias current. In order to obtain the minimum error, the chosen bias current should be in the flat region of the β curve [4.10], such that $\beta(I) \sim \beta(pI)$. Due to process spread, however, this flat region may shift, and in advanced technology nodes such a region may not even exist [4.9]. In such cases, this β-compensation method would not be very effective, as it only compensates for the effect of β on V_{BE} (second term in 3.6), but not for the effect of β on ΔV_{BE} (second term in 3.8).

Fig. 4.5 (a) shows an ideal scenario, where β is compensated in both V_{BE} and ΔV_{BE}. This involves copying the base current of Q_L, and Q_R and then adding them to their own emitter currents. This, in turn, forces the collector currents of Q_R and Q_L to be purely defined by I_{bias} and $5 \cdot I_{bias}$, which is required for an ideal bipolar core. As illustrated in Fig. 4.6 (b), the trimmed temperature spread of Fig. 4.4 (b) is then reduced by a further 2× to ±0.03 °C.

The trimmed results in Fig. 4.4 (a), and also in Fig. 4.6 (a), exhibit a systematic non-linearity, which originates from the inherent non-linearity of V_{BE}. Optimizing the value of α could significantly reduce this. It has been shown that choosing α that makes V_{REF} $(= V_{BE} + \alpha \cdot \Delta V_{BE})$ slightly PTAT helps to reduce this non-linearity [4.1]. When it is necessary to further suppress this non-linearity, a higher-

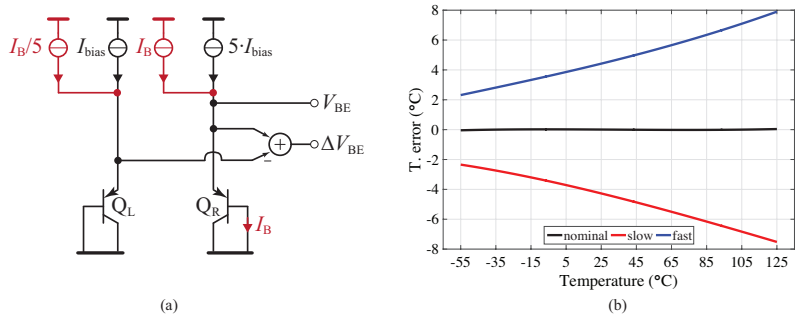


Figure 4.3: (a) Bipolar core with β compensation in V_{BE} and without R_S compensation. (b) Untrimmed temperature error.

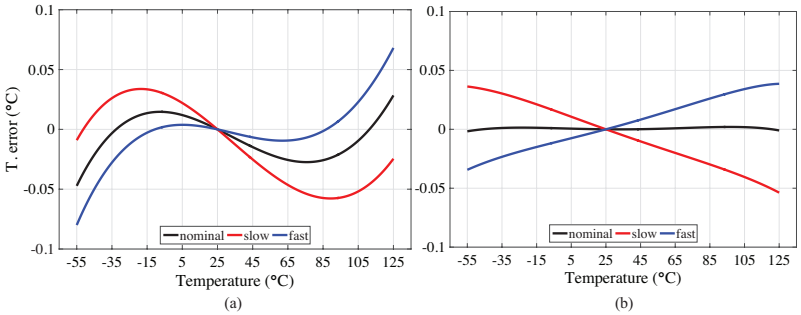


Figure 4.4: Trimmed temperature error for the bipolar core of Fig. 4.3 (a) with a 1st-order polynomial fit and (b) with a 3rd-order polynomial fit.

order polynomial can be used to map μ_T to D_{out}. In case of a third-order polynomial, $D_{out} = a_0 + a_1 \cdot \mu_T + a_2 \cdot \mu_T^2 + a_3 \cdot \mu_T^3$, where $a_0 : a_3$ are constants. In Fig. 4.4 the use of a third-order polynomial fit does not improve accuracy much, due to the relatively large errors caused by β spread. The advantage of using the third-order polynomial become more obvious in Fig. 4.6 (and also in Fig. 4.8), in which β-related errors are sufficiently suppressed. The remaining residual error in this figure is then mainly due to the uncompensated effect of R_S.

4.2.3 Compensated β and R_S

R_S-related errors in V_{BE} (the second term in (3.10)) are proportional to the value of bias current as well as to the value of R_S. One way to suppress them is by simply choosing a sufficiently small bias current. However, the level of bias current cannot always be freely chosen, since it must also satisfy the requirement that $\beta(I) \sim \beta(pI)$, or it must be enough to obtain a desired g_m in the PNPs for settling purposes. R_S-

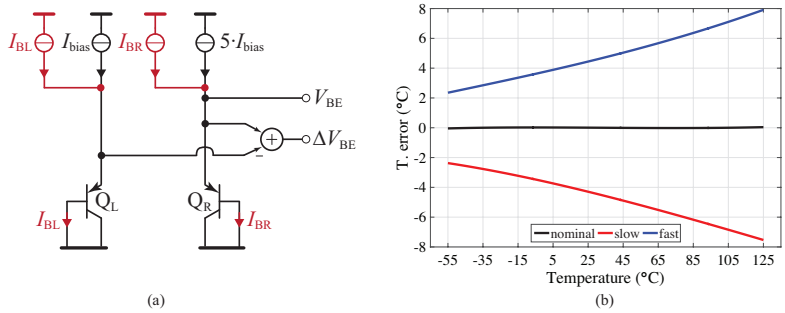


Figure 4.5: (a) Bipolar core with β compensation in both V_{BE} and ΔV_{BE} and without R_S compensation. (b) Untrimmed temperature error.

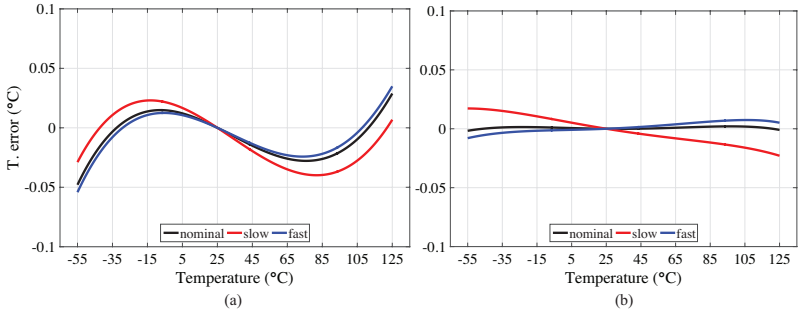


Figure 4.6: Trimmed temperature error for the bipolar core of Fig. 4.5 (a) with a 1st-order polynomial fit and (b) with a 3rd-order polynomial fit.

related errors can also be removed by noting that it is a linear function of I_{bias}, while V_{BE} is a logarithmic function of I_{bias} [4.12]. With reference to (3.10), and assuming ideal β-compensation, the effect of R_S can be removed by using the following function:

$$\begin{aligned} V_{BE}(I_{bias}) &= 2 \cdot V_{BE}(2 \cdot I_{bias}) - V_{BE}(4 \cdot I_{bias}) \\ &= \frac{kT}{q} \cdot ln(\frac{I_{bias}}{I_S}) + 2 \cdot R_S \cdot (2 \cdot I_{bias}) - R_S \cdot (4 \cdot I_{bias}) \,. \end{aligned} \tag{4.2}$$

Fig. 4.7 (a) illustrates a bipolar core in which this R_S-compensation scheme is used for both V_{BE1}, and V_{BE2}. Together with β-compensation, this results in a trimmed inaccuracy of ± 0.003 °C, as illustrated in Fig. 4.8 (b). This negligible error originates from the slight dependency of R_S on β, which makes it slightly current-dependent. As a result, the R_S compensation described in (4.2) is not perfect.

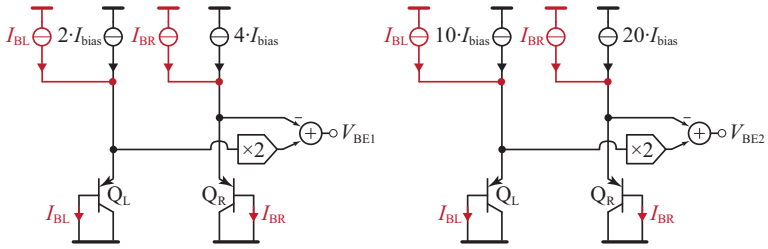


Figure 4.7: Bipolar core with β and R_S compensation for both V_{BE1} and V_{BE2}.

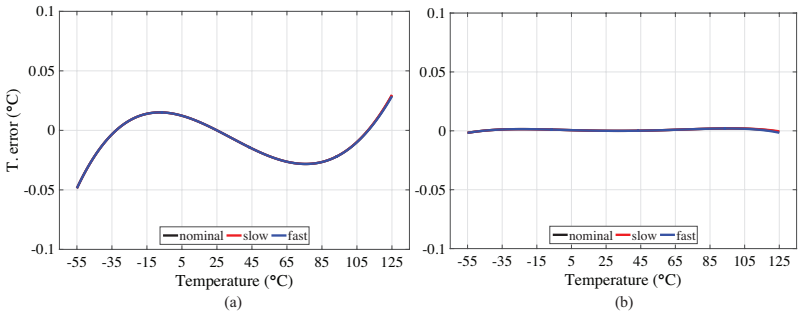


Figure 4.8: Trimmed temperature error for the bipolar core of Fig. 4.7 (a) with a 1st-order polynomial fit and (b) with a 3rd-order polynomial fit.

4.3 Inaccuracy sources in the bias circuit

As illustrated in Fig. 4.9, the bias circuit generates a PTAT current. Compared to a constant current, a PTAT bias current reduces the nonlinearity in V_{BE} [4.13]. The bias circuit uses two PNPs at a current density ratio of $1 : p_{Eb}$. An opamp then forces the resulting ΔV_{BEb} ($= V_{BEb2} - V_{BEb1}$) across an n+ poly bias resistor, R_b. The bias current $I_{bias} = \Delta V_{BEb}/R_b$ can then be copied to the bipolar core with a current ratio of $1 : m$. Setting $m = 1$, and $p_{Eb} = p_E = 5$ simplifies the dynamic techniques used for the matching of current mirrors (see Sec. 4.4.1). $R_b = 500$ kΩ sets the value of I_{bias} such that $\beta(I) \sim \beta(p_{Eb} \cdot I)$. Adding a second n+ poly resistor $R_b = R_b/p_{Eb}$ makes the bias current become slightly β-dependent [4.1],[4.11], i.e., $I'_{bias} = I_{bias} \cdot (\beta + 1)/\beta$, where β is the current gain of Q_{Rb}. The β-dependent I'_{bias}, can then cancel the effect of β on the bipolar core's V_{BE} ($= V_{BE2}$), assuming that β in Q_R and Q_{Rb} are equal. This is equivalent to the scenario described in Fig. 4.3. As a result, V_{BE} can be

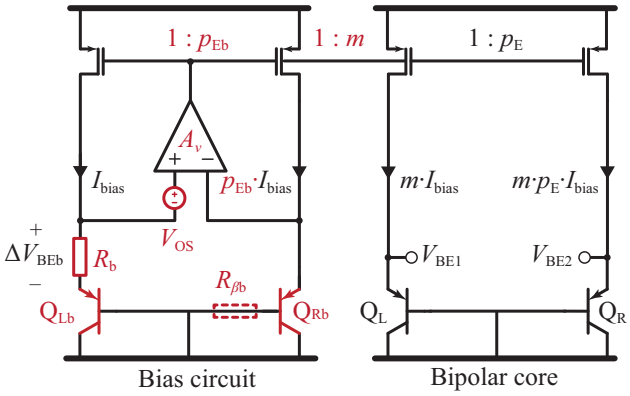


Figure 4.9: Bias circuit that generates a PTAT current. Sources of error are shown in red.

expressed as:

$$V_{BE} = \frac{kT}{q} \cdot ln(\frac{I'_{bias}}{I_S} \cdot \frac{\beta}{\beta+1})$$
$$= \frac{kT}{q} \cdot ln(\frac{I_{bias}}{I_S} \cdot \frac{\beta+1}{\beta} \cdot \frac{\beta}{\beta+1}) = \frac{kT}{q} \cdot ln(\frac{I_{bias}}{I_S}) . \quad (4.3)$$

Different sources of error in the bias circuit can cause I_{bias} to deviate from its nominal value, which then causes errors in V_{BE} and in D_{out}. If these errors only change the absolute value of I_{bias} (i.e., $\Delta I_{bias}/I_{bias} = \epsilon$), they create an error similar to that of I_S in (4.1). As a result, this error will be corrected for by the same PTAT trim that corrects for the spread of I_S. However, if such errors change the temperature dependence of I_{bias}, they will cause residual temperature errors after the trim.

The main source of error in I_{bias} is the spread of the bias resistor R_b. The absolute value of R_b substantially spreads, but this will be completely removed by the PTAT trim since it only changes the absolute value of I_{bias}. Spread in the temperature dependence of R_b, however, results in residual error after the trim. To obtain high accuracy, it is therefore essential that the temperature dependence of R_b is minimal. In the 0.16-μm technology, a non-silicided n+ poly resistor is the most suitable candidate for this reason. As shown in Fig. 4.10, over the process corners, the absolute value of R_b varies by $\pm 18\%$, while its temperature coefficient varies by a maximum of $\pm 0.01\%/°C$. The former creates a large untrimmed error of 2.18 °C but no trimmed error, while the latter causes ± 0.11 °C error after the trim.

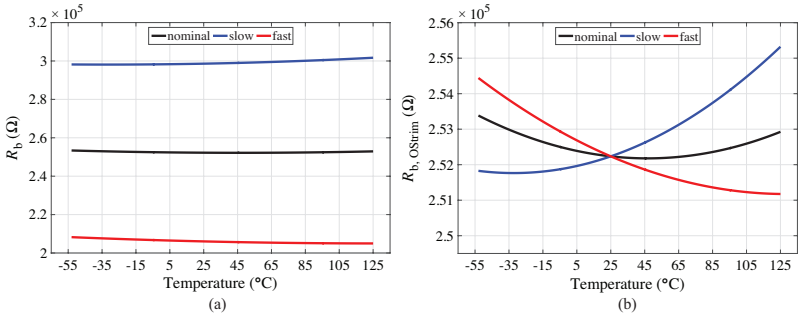


Figure 4.10: (a) Variation in n+ poly resistance in the 0.16-μm technology. (b) Variation in temperature dependence of the resistance.

Limited gain and non-zero offset of the opamp can also cause I_{bias} to change. With reference to Fig. 4.9, when the opamp has a limited gain of A_v, I_b can be written as follows:

$$I_{bias} \approx \frac{\frac{kT}{q} \cdot ln(p_{Eb})}{R_b} \cdot (1 - \frac{1}{G_m \cdot A_v \cdot R_b}), \tag{4.4}$$

where G_m is the transconductance of the current mirrors. Assuming $G_m = 1$ μA/V and $R_b = 500$ kΩ, a 60-dB gain causes a 0.6 nA variation in I_{bias} compared to an infinite gain. In practice, however, only the spread in A_v appears as temperature sensing error. The opamp's non-zero offset V_{OS} can also cause I_{bias} to vary by $\Delta I_{bias} = V_{OS}/R_b$. Assuming a 2 mV temperature-independent offset for the opamp, I_{bias} changes by 9 nA, which then creates a maximum of 0.4 °C untrimmed error. After the trim, this error reduces to 0.14 °C.

Similar to the PNPs in the bipolar core, Q_{Rb} and Q_{Lb} in the bias circuit have limited β and non-zero R_S, which can cause error in I_{bias}. Variation in β and R_S only affect ΔV_{BEb}, which has a much lower influence on the temperature sensor accuracy than ΔV_{BE}. Furthermore, some of these errors will be suppressed or completely removed by the PTAT trim. The temperature sensing error caused by variation of β, either from its absolute value β_0 or its temperature dependence XTB, is shown in Table 4.1. For these calculations, variations of β_0 and XTB are obtained from the modeled parameters of the PNPs, and also $\Delta\beta = 0.1$ is assumed as the difference between the β of Q_{Rb} and Q_{Lb}. R_S can also cause error in I_{bias} in the same way as R_b. As a result, variations in the absolute value of R_S will be completely removed after the trim but variations of its temperature dependence will not. Compared to R_b, the relatively small value of $R_S \sim 200$ Ω creates much lower temperature-sensing error.

The last source of error in I_{bias}, but certainly not the least, is the variation of the current ratios $1 : p_{Eb}$ and $1 : m$. In practice, these ratios rely on the matching of the PMOS transistors used to realize the current mirrors. When biased in strong inversion, the error in the current I of two identical current mirrors can be expressed as:

$$\frac{\Delta I}{I} \approx \frac{2}{\frac{V_{gs}}{V_T} - 1} \cdot |\frac{\Delta V_T}{V_T}| + |\frac{\Delta \mu_p}{\mu_p}| + |\frac{\Delta C_{ox}}{C_{ox}}| + |\frac{\Delta W}{W}| + |\frac{\Delta L}{L}|\,, \tag{4.5}$$

where μ_p is the mobility of holes, C_{ox} is the gate-oxide capacitance per unit area, W and L are the transistor dimensions, V_T is the threshold voltage, and V_{gs} is the gate-source voltage of the current mirrors. As described in [4.14], the parameters of (4.5) have a statistical nature, and thus the variation of I follows:

$$\frac{\sigma^2(I)}{I^2} \approx \frac{4 \cdot \sigma^2(V_T)}{(V_{gs} - V_T)^2} + \frac{\sigma^2(\mu_p)}{\mu_p^2} + \frac{\sigma^2(C_{ox})}{C_{ox}^2} + \frac{\sigma^2(W)}{W^2} + \frac{\sigma^2(L)}{L^2}\,, \tag{4.6}$$

where σ represents the standard deviation of each parameter. The variation of each parameter of (4.5) should be modeled with the corresponding mismatch-generating processes. For V_T, and similarly for μ_p and C_{ox}, the mismatch between two devices at a distance D on a same wafer follows a general rule [4.14]: $\sigma^2(V_T) = A_{V_T}^2/(WL) + S_{V_T}^2 \cdot D^2$, where A_{V_T} is the area proportionality constant, and S_{V_T} is the constant that describes the variation with distance. For W, and equally for L, the mismatch generating process is slightly different than for V_T, since it originates from the edge roughness of the gate which then causes $\sigma^2(W) \propto 1/L$. As a result, its contribution to (4.6) can be expressed as: $A_W^2/L^2W + S_W^2 D^2$. It is worth mentioning that the D for the two current mirrors can be assumed to be zero.

The main observation from (4.5) and (4.6) is that $1 : p_{Eb}$ is not necessarily temperature independent, since it is a function of many temperature dependent parameters including V_T and V_{gs}. In the target 0.16-μm technology, Monte-Carlo simulations on identical current mirrors showed that its absolute value changes by $\pm 15\%$, while its temperature dependence also varies by $\pm 0.055\%/°\text{C}$. The 15% variation in the absolute value of p_{Eb} causes a 1.17 °C untrimmed error, which will be removed by the PTAT trim. The $\pm 0.055\%/°\text{C}$ variation in the temperature dependence of p_{EB}, and similarly for m, results in a ± 0.36 °C error after the trim.

Table 4.1 summarizes the various temperature-sensing errors that originate from the bias circuit. These results are based on the available modeled parameters for R_b, β, p_{Eb} which are themselves only approximations of reality. For the spread of A_v, V_{OS}, and temperature dependence of R_S, reasonable assumptions were made.

Item		Nominal	Variation	Untrimmed (°C)	Trimmed (°C)
R_b	value	500 $k\Omega$	±18%	±2.18	0
	TC	0	±0.01%/°C	0.11	0.11
P_{Eb}	value	5	±15%	1.17	0
	TC	0	±0.05%/°C	0.36	0.36
V_{OS}	value	0	±2 mV	0.4	0.14
A_v	value	60 dB	±20 dB	0.002	0
β	β_0	4.5	±20%	0.009	0.006
	XTB	2	±0.5	0.003	0.003
R_S	value	200 Ω	±30%	0.007	0
	TC	0	±1%/°C	0.02	0.02

Table 4.1: Worst-case temperature-sensing error originating from the bias circuit from −55 °C to 125 °C . Variations are based on MonteCarlo simulations of model parameters in the 0.16-μm (for R_b, β, p_{Eb} and the value of R_S) or a reasonable assumption (for A_V, V_{OS}, and TC of R_S).

Based on the simulated accuracy of the bipolar core in Sec. 4.2, and the error sources in the bias circuit shown in Table 4.1, some important conclusions can be drawn about the accuracy of BJT-based TDCs.

(1) Fixed calibration parameters are necessary for low-cost calibration.

(2) The untrimmed accuracy of BJT-based TDCs is very poor due to the large spread of I_S and R_b.

(3) BJT-based TDCs can be very accurate after a PTAT trim.

(4) Obtaining high accuracy after trimming requires a bias current I_{bias}, with a well-defined temperature dependence.

4.4 Sensing front-end

Fig. 4.11 shows the sensing front-end of the TDC. It is essentially similar to Fig. 4.9 with two extra features: (1) dynamic element matching is used to correct errors in I_{bias}, and (2) the PNP's current gain can be detected. Similar to [4.1],[4.2], the PNPs in the bias circuit and the bipolar core are biased at 1 : 5 current ratios. The unit bias current is 80 nA, a choice that ensures that the PNPs' are biased in their flat-β region. The resulting V_{BEb} is forced across an n+ poly resistor, R_b, to generate I_{bias}. A second n+ poly resistor, $R_{\beta b}(= R_b/5)$, is then used to cancel β in the V_{BE} of the bipolar core. This β-compensation relies on the matching of Q_R and Q_{Rb} and also on the 1 : 5 ratios of $1 : p_{Eb}$, and $R_{\beta b} : R_b$.

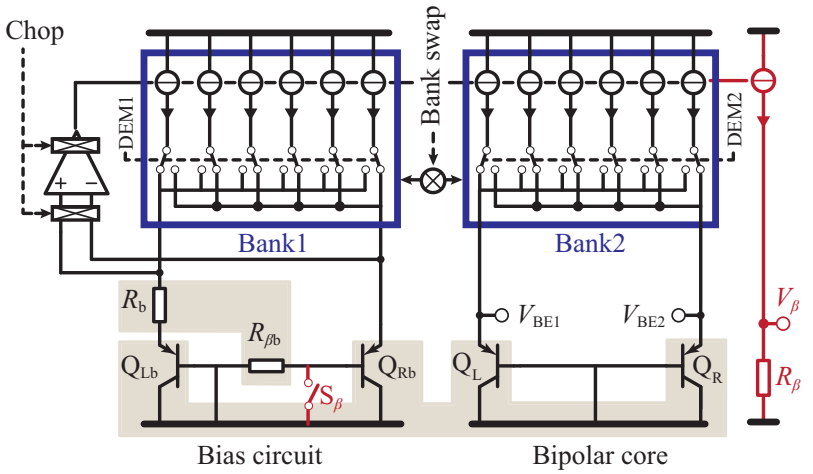


Figure 4.11: Sensing front-end with dynamic techniques to suppress the sources of error. (Red) β-detection circuitry. (Gray) careful layout for matching purposes.

4.4.1 Bias circuit

To ensure high accuracy, the sources of error that affect the temperature dependence of I_{bias} must be canceled out. These errors mainly originate from the temperature dependence of R_b, the opamp's offset, and the p_{EB} and m ratios. A single-point trim would then correct for variations in the absolute value of I_{bias}. To reduce the opamp's related error, a folded-cascode structure is adapted with a minimum gain of 80 dB, with less than 10 dB variation over process and temperature. The opamp's offset is also chopped using the chopper switches at its input and output. Similar to [4.5] and [4.15], to mitigate the errors caused by current ratios, the current mirrors that realize the bias circuit $1 : p_{EB}$ ratio are dynamically matched with the help of the control signal denoted by DEM1 in Fig. 4.11. In the same fashion, the current mirrors that realizes $1 : p_E$ in the bipolar core are matched using DEM2. Furthermore, to ensure that the current is accurately copied from the bias circuit into the bipolar core (i.e., $m = 1$), the two banks of current mirrors are dynamically swapped (Bank-swap). Since the two banks are designed to have the same current levels, this is accomplished by adding extra switches in series with each current bank. The accuracy of the ratio $R_{\beta b} : R_b$ and the matching of Q_R and Q_{Rb} rely on careful layout.

4.4.2 β detection

The sensing front-end is capable of determining β directly, which then allows its effect on the accuracy of the TDC to be observed experimentally. As in [4.1], [4.11], the effect of β in V_{BE} is compensated by using $R_{\beta b}$. This, however, does not correct the effect of β on ΔV_{BE}. This increases the TDC's error slightly and influences the accuracy of voltage calibration. Detecting β can then be used as a means of determining the origin of the measured spread of ΔV_{BE}.

The β-detection circuit in the sensing front-end measures the β of Q_{Rb}. Via switch S_β the $R_{\beta b}$ can be turned off or on, which in turn changes the bias current from I_b ($= V_{BEb}/R_b$) to I'_b ($= V_{BEb}/R_b \cdot (\beta+1)/\beta$). The ratio of these two currents $X_\beta = I_{bias}/I'_{bias}$ is then used to extract the value of β. To obtain X_β, the TDC uses a sense resistor R_β ($= 350$ kΩ) whose voltage drop V_β is digitized by the ADC.

β-detection involves two steps. First, switch S_β is off and the ADC digitizes $X_{1\beta}$ as the ratio between $V_{1\beta}$ ($= R_\beta \cdot n \cdot I_{bias}$) and ΔV_{BE1}. Second, switch S_β is on, and the ADC digitizes $X_{2\beta}$ as the ratio between $V_{2\beta}$ ($= R_\beta \cdot n \cdot I'_{bias}$) and ΔV_{BE2}. In the digital backend, β is computed using $\beta = X_{2\beta}/(X_{1\beta} - X_{2\beta})$. Since β-detection is a ratiometric measurement, it is not affected by the gain errors of the current copying ratio n, or the variation in R_β, due to process or temperature variations. The two conversions required in β-detection occur in quick succession (<200 ms), and so $(V_{1\beta}, \Delta V_{BE1})$ and $(V_{2\beta}, \Delta V_{BE2})$ are assumed to be at the same die temperature T_{die}. ΔV_{BE1} and ΔV_{BE2} can be assumed to be identical as well, since at the two states of S_β, the change in the absolute value of I_{bias} barely changes ΔV_{BE}. Simulations showed that I_{bias} changes by 12% at 25 °C and 25% at 125 °C, which correspond to a 0.002% and 0.04% change in ΔV_{BE}.

4.5 Readout circuit

An overview of the TDC is shown in Fig. 4.12. It consists of an analog core and a digital backend. The analog core is composed of the sensing front-end which produces V_{BE}, ΔV_{BE}, and V_β, and the zoom-ADC, which digitizes the ratio of voltages. In normal operation mode, the decimated output of the zoom-ADC provides $X_T = V_{BE}/\Delta V_{BE}$. The TDC can also be configured to digitize $X_\beta = V_\beta/\Delta V_{BE}$ in β-detection mode, or $X_V = V_{ext}/\Delta V_{BE}$ in the voltage calibration mode, in which V_{ext} is an external voltage.

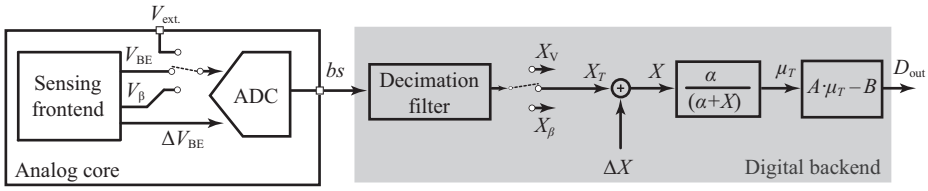


Figure 4.12: Overview of the TDC. The analog core is realized on-chip, while the digital backend is realized off-chip.

4.5.1 Zoom-ADC

A zoom-ADC obtains high energy efficiency by combining the benefits of two ADC architectures [4.2],[4.16]–[4.18]. It is based on a ΣΔ modulator whose references are zoomed in around the input signal, instead of at the full-scale levels in conventional modulators. This, in turn, reduces the amount of error signal that is fed into the ΣΔ's loop-filter and thus relaxes the requirement for its first active integrator. When the input signal changes, the levels of reference signals also need to change to bound the input signal with enough margin and ensure a stable modulator. In the zoom-ADC, this is obtained with an energy-efficient 6-bit SAR conversion, which coarsely digitizes the input signal. The ΣΔ modulator then selects the right level for the reference signals based on this coarse conversion.

Given the quasi-static nature of temperature, a static zoom-ADC similar to [4.2] has been adapted, where an initial SAR conversion is followed by a 2nd-order ΣΔ conversion. To save area, the SAR and ΣΔ conversions share the same circuitry, as shown in Fig. 4.13. In the first step, the SAR algorithm is realized by using the first stage of the modulator (as a charge-amplifier), the comparator, and the 31-element capacitive DAC (C-DAC). The 6-bit SAR determines the integer part of X_T (i.e. K_{SAR}) by successively comparing V_{BE} with $K \cdot \Delta V_{BE}$, where K $(= 1 : 31)$, is realized with the C-DAC. In the second step, the circuit is configured as a 2nd-order modulator, which balances V_{BE} against the reference voltages $(K_{SAR}-1)\cdot\Delta V_{BE}$ and $(K_{SAR}+1)\cdot\Delta V_{BE}$. This results in a bitstream bs with an average that is equal to $\mu_{\Sigma\Delta} = [V_{BE} - (K_{SAR}-1)\cdot\Delta V_{BE}]/(2\Delta V_{BE})$. The two outputs of SAR and ΣΔ are then combined in the digital domain to obtain X_T by using $X_T = (K_{SAR}-1)+2\cdot\mu_{\Sigma\Delta}$.

Dynamic techniques ensure that the zoom-ADC's gain and offset errors are adequately suppressed. As in [4.19], Correlated Double Sampling (CDS) is used in the first stage of the ΣΔ to suppress the effect of the offset and $1/f$ noise of the first amplifier. Any residual offset is then further suppressed by using system-level chopping that occurs once per conversion. A data-weighted averaging DEM [4.20] is employed

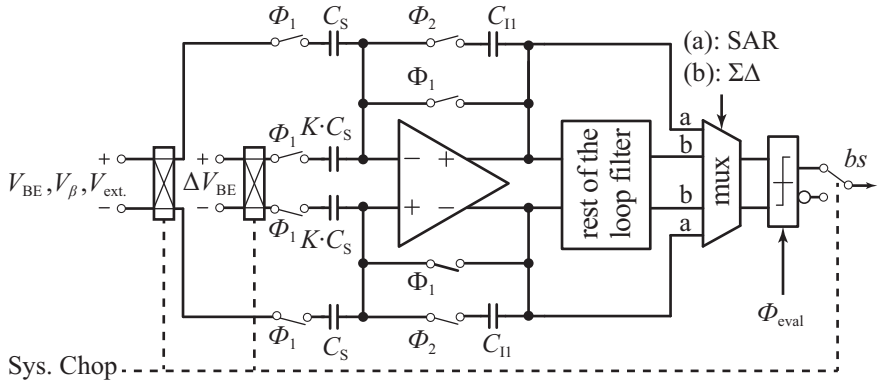


Figure 4.13: Zoom-ADC. It is composed of a SAR and a 2nd-order $\Delta\Sigma$ ADC using a shared circuitry.

to remove the mismatch in the C-DAC elements to reduce the ADC's gain errors. This involves shuffling the position of the capacitor that samples V_{BE}, together with the K capacitors that sample ΔV_{BE} at the rate of the sampling rate f_s.

In parallel to C-DAC DEM, and in the same fashion, the unit currents in the bipolar core (Bank2) and in the bias circuit (Bank1) are dynamically matched. Therefore, a unit current biasing Q_L (or Q_{Lb}) and 5× unit currents biasing Q_R (or Q_{Rb}) shuffle around a 6-element current DAC at the rate of f_s. Bank Swapping, on the other hand, occurs at the rate of $f_s/5$, and so after a complete DEM cycle for Bank1 and Bank2. Chopping Q_R and Q_L occurs at the rate of $2f_s$, and thus, between ϕ_1 and ϕ_2 their positions swap.

Each of the two integrators is built around an energy-efficient current-reuse amplifier as shown in Fig. 4.14. This amplifier uses both PMOSFET and NMOSFET input devices to obtain a higher equivalent g_m $(= g_{mN} + g_{mP})$ than the telescopic amplifier used in [4.2]. Inverter-based amplifiers [4.21] can also obtain a higher equivalent g_m $(= g_{mN} + g_{mP})$, but they require dynamic biasing to set the inverter's output common-mode level. In addition, they need to operate in a pseudo-differential fashion with two single-ended branches that are fed with differential signals. Compared to an inverter-based amplifier, a current-reuse amplifier is less complex and avoids the additional sampled noise associated with dynamic biasing. Furthermore, it is a fully differential circuit, which results in higher supply and common-mode rejection.

The first amplifier consumes 480 nA at RT, while the second amplifier is scaled down by 4×, drawing 120 nA. In the current-reuse structure, the DC voltage level of the two inputs are inevitably equal, and it is best to set their common-mode voltage at

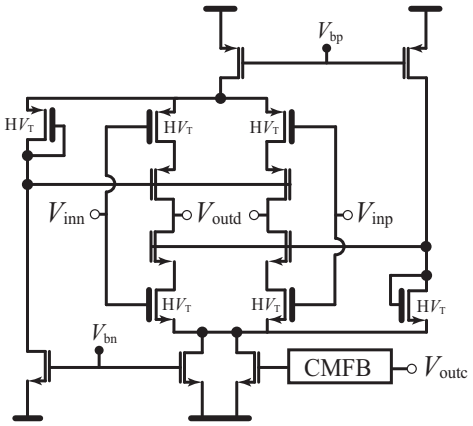


Figure 4.14: Current-reuse amplifier used in the first and the second stage of the zoom-ADC.

$V_{dd}/2$. In this structure, however, the swing of the amplifier is limited to $V_{TN} + V_{TP}$, where V_T is the threshold voltage. To obtain a higher swing, the input devices have been realized with available high threshold voltage (HV_T) devices.

4.5.2 Switch leakage

A simplified diagram of the zoom-ADC's sampling network is shown in Fig. 4.15. As shown, the leakage currents (I_{L1} and I_{L2}) of the sampling switches alter the PNP's bias current and thus cause errors in V_{BE}. Furthermore, they flow through the ON-resistance, R_{ON}, of the switches and cause errors in the sampled voltage V_S. There are two components of the leakage current: I_{L1} and I_{L2}. The former is associated with the reversed-biased junction between the substrate and the source and drain of the sampling switches, while the latter is the current that flows through the OFF-resistance, R_{OFF}, of any other switches connected to C_S. The sampling error due to I_{L1} is directly proportional to the total number of switches connected to the PNPs, while the sampling errors due to I_{L2} are a function of $R_{ON}/(N \cdot R_{OFF})$, where N is the number of OFF switches involved.

In the zoom-ADC, the sampling network consists of a capacitor array with 64-unit elements (in differential branches), each of which can connect to V_{BE1}, V_{BE2} and V_{ext} via one of three sampling switches. These switches enable the application of DEM to the capacitor array, which is necessary to mitigate the effect of capacitor mismatch but results in a total of 192 sampling switches. This, in turn, can cause

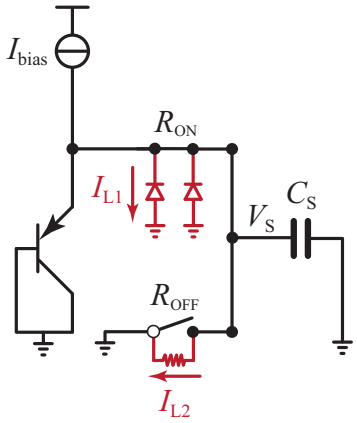


Figure 4.15: Switch leakage in the sampling network of the zoom-ADC.

sampling errors, particularly at higher temperatures, where the leakage currents are largest. Different TDC versions, using different switches, were developed to explore the impact of leakage currents.

4.6 Experimental results

The TDC is realized in 0.16-μm CMOS technology. The analog core occupies a silicon area of 0.16 mm^2, as shown in Fig. 4.16, and draws 4.6 μA of current from a nominal 1.8 V supply. A $sinc^2$-decimation filter, mapping and trimming functions, the SAR and DEM logic are realized off-chip for flexibility. Table 4.2 summarizes the TDC's current and area consumption. In this table, the numbers related to the off-chip digital circuitry are based on post-synthesis realizations.

A key performance metric of a TDC is its output noise, which is usually designed to be an order of magnitude lower than the target accuracy. Fig 4.17 (a) illustrates the output *rms*-noise versus the conversion time. Given the quasi-static nature of temperature and the use of the $\Sigma\Delta$-based readout, the noise and conversion time can be exchanged arbitrarily. In the normal operation mode, with a 35-kHz sampling clock frequency, the TDC obtains a resolution of 15 mK with a conversion time of 5 ms. During calibration and trimming, when a higher resolution is required, the conversion time is increased to 100 ms, in which case the TDC obtains a resolution of 4 mK.

The TDC's power supply sensitivity is another important metric. At RT, its output changes by 10 mK when the supply voltage varies from 1.5 to 2 V (Fig.

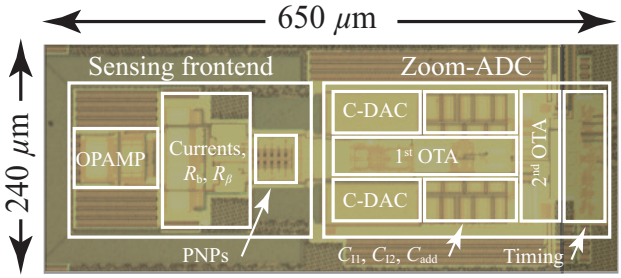


Figure 4.16: Die micrograph.

	Current (μA)	**Area** (mm^2)
Sensing front-end	3.4	0.063
Zoom-ADC	1.2	0.093
Digital controller (SAR, DEM)	1.12	0.046
Decimation and calibration	0.19	0.163

Table 4.2: Breakdown of current and area consumptions. The current consumptions are for 25°C.

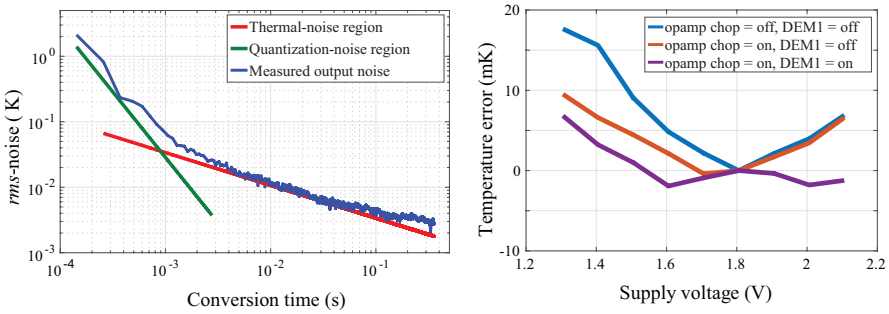


Figure 4.17: (a) Measured output noise versus conversion time. (b) Measured supply sensitivity in different modes of the bias circuit.

4.17 (b)), which corresponds to a 0.01°C/V supply sensitivity. This illustrates the effectiveness of the dynamic techniques used in the front-end, showing how the supply sensitivity improves as the various dynamic techniques are activated.

4.6.1 Batch-calibrated accuracy

Within one production batch, variations in device parameters are quite limited. The effect of mismatch and the effectiveness of dynamic techniques to suppress them can be investigated by observing sample-to-sample variations. Twenty samples of the TDC taken from one wafer were characterized in a climate chamber from −55 °C to 125 °C. The TDCs were packaged in ceramic to minimize the stress-related errors on the TDC dies. The packaged samples were then mounted on a metal block, whose temperature T_{REF} was accurately measured, with less than ±1 mK error, by a calibrated PT-100 thermistor. Outputs of this thermistor and the TDCs were then captured at multiple temperature points to obtain the temperature error of the TDCs.

The average characteristic of these 20 samples was used to extract the average parameters, A, B, and α—a procedure known as batch calibration. The average of the ADC's output X_T led to a master curve $X_{T,avg}$. α was then found by minimizing the non-linearity in $\mu_{T,avg} = \alpha/(\alpha + X_{T,avg})$. This involved minimizing the Least-Square error between $\mu_{T,avg}$ and a straight line. In a similar fashion, values of A, and B were found by minimizing the Least-Square error between $D_{out,avg} = A \cdot \mu_{T,avg} - B$ and T_{REF}. These batch-calibrated parameters were then applied to all 20 TDC samples of the batch. Fig. 4.18 (a) illustrates the inaccuracy of the TDC without trimming, which is ±0.4 °C (±3σ).

The main source of this error is the spread of I_S as well as the spread of R_b. These variations can be corrected for by calibrating the sample's V_{BE} and then trimming it in a PTAT manner. Alternatively, the ADC's output X_T can be calibrated at 25°C and then digitally trimmed by adding a fixed offset X_{OS} as:

$$X_{T,trim}(T) = X_{T,untrim}(T) + X_{OS}\,, \tag{4.7}$$

$$X_{OS} = X_{T,avg}(25°C) - X_{T,untrim}(25°C)\,, \tag{4.8}$$

where $X_{T,trim}$ and $X_{T,untrim}$ are the ADC's trimmed and untrimmed outputs. The TDC's trimmed output is then calculated as:

$$D_{out,trim}(T) = A \cdot \frac{\alpha}{\alpha + X_{T,trim}(T)} - B\,. \tag{4.9}$$

Fig. 4.18 (b) shows the temperature error after trimming, which is ±0.1 °C (±3σ). The residual nonlinearity of the TDC is about 40 mK, which is considerable compared to the total temperature error. The residual non-linearity can be further reduced to

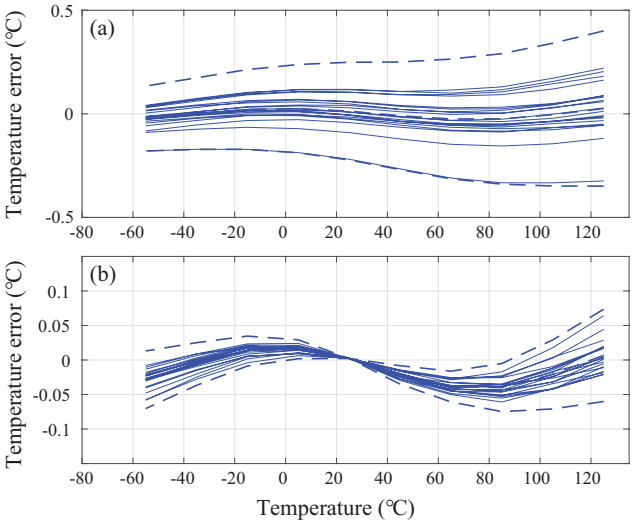


Figure 4.18: Batch-calibrated inaccuracy using three fitting parameters A, B, and α: (a) untrimmed and (b) trimmed. Dashed lines represent the average and $\pm 3\sigma$ boundaries.

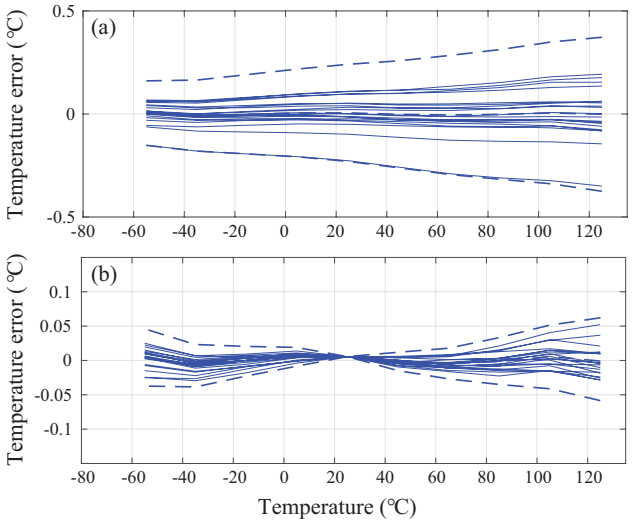


Figure 4.19: Batch-calibrated inaccuracy using a 3rd-order polynomial fit: (a) untrimmed and (b) trimmed. Dashed lines represent the average and $\pm 3\sigma$ boundaries.

less than 10 mK by using a 3rd-order polynomial fit. To illustrate this, results shown in Fig. 4.18 have been replicated in Fig. 4.19 with a 3rd-order polynomial fit. The coefficients for the polynomial have been found to provide the Least-Square error between $D_{out,avg}$ and T_{REF}. Using this polynomial, the TDC obtains an untrimmed

Opamp-chop	DEM-1	Bank-swap	Untrimmed error	Trimmed error
off	off	off	$\pm 700\ mK$	$\pm 300\ mK$
on	off	off	$\pm 400\ mK$	$\pm 100\ mK$
on	on	off	$\pm 380\ mK$	$\pm 70\ mK$
on	on	on	$\pm 380\ mK$	$\pm 60\ mK$

Table 4.3: Accuracy improvement by applying dynamic techniques in the bias circuit. The reported result uses the 3rd-order polynomial fit.

	This work	[4.1]	[4.2]	[4.22]	[4.23]
CMOS technology (μm)	**0.16**	0.7	0.16	0.065	0.022
Area (mm^2)	**0.16**	4.5	0.08	0.2	0.004
Supply voltage (V)	**1.5-2**	2.5-5.5	1.5-2	1.5	1-1.3
Current consumption (μA)	**4.6**	75	3.4	0.5	50
Temperature range ($^\circ C$)	**−55-125**	−55-125	−55-125	−40-130	−30-120
3σ-inaccuracy (mK)	**±60**[a]	±100	±150	±1070	±150
Power supply sensitivity ($^\circ C/V$)	**0.01**	0.03	0.5	N/A	N/A
Resolution (mK)	**15**	10	20	125	580
Conversion time (ms)	**5**	100	5.3	2	0.032
Resolution FOM[b] (pJK^2)	**7.8**	1875	11	23	540
Relative inaccuracy[c] (%)	**0.07**	0.11	0.17	0.47	1.4

[a] Batch-calibrated result with a 3rd-order fit
[b] Resolution FOM = Power×Conversion time×Resolution^2
[c] Relative inaccuracy = 2×Inaccuracy×100/Temperature range

Table 4.4: Performance summary and comparison with other accurate CMOS temperature sensors.

inaccuracy of ±380 mK (±3σ), and a trimmed inaccuracy of ±60 mK (±3σ).

The remaining error of the TDC after removing the TDC's systematic non-linearity and I_S-related error partly stems from non-PTAT error sources in the bias circuit. To observe this, the trimmed inaccuracy of the TDC was obtained using different sets of dynamic techniques in the bias circuit. Table 4.3 summarizes these results, and shows that the inaccuracy improved by gradually applying these techniques.

Table 4.4 summarizes the TDC's performance within a batch and compares it with other related work. The designed TDC achieves superior performance in terms of inaccuracy, and supply sensitivity, while it obtains a comparable energy efficiency FOM.

	$\mathbf{UHV_T}$	$\mathbf{HV_T}$	$\mathbf{V_T}$
Batch-1	20 samples	20 samples	-
Batch-2	-	20 samples	20 samples
Batch-3	20 samples	-	-

Table 4.5: Summary of the implemented batches and technology options of the TDC.

4.6.2 Batch-to-batch variation

The spread of device parameters is wider over multiple batches than within one batch, which might result in a larger temperature error. To evaluate the TDC's inaccuracy over a larger spread, three different batches were fabricated and then characterized. Each batch was taken from a different fabrication run with a time difference of a few months, thus they were expected to be somewhat different from each other. These batches, however, do not represent the corners of the technology.

The TDC of each batch also used the available technology flavors to observe the effect of switch leakage on the ADC's sampling circuit. The ADC's sampling switches were implemented with different threshold voltages, and so with different switch OFF-resistance. Table 4.5 summarizes the implemented batches and flavors of the TDC. UHV_T (Ultra High Threshold Voltage) devices were used in the TDCs of Batch-1, and Batch-3. These devices have the largest V_T compared to normal V_T devices and thus have the lowest leakage. However, they require extra fabrication masks for realization, making them more expensive. To avoid the higher costs, HV_T (High Threshold Voltage) devices were also used in Batch-1 and Batch-2. These devices have a larger V_T than normal devices, but lower than UHV_T devices.

The measured results of Batch-1 and Batch-2 are shown in Fig. 4.20. Average parameters were extracted from Batch-1 and then used for both batches. This way, the extra batch calibration that was required in [4.2] was avoided, which significantly saved cost and simplified the calibration process. Without trimming, the TDCs' output spread by 2 °C. A single-point trim, however, reduced this error to less than ±0.1 °C (3σ) and maintained the batch-calibrated accuracy.

Using the β-detection circuit, the β for each sample of the two batches was measured and compared. Fig. 4.20 (c) illustrates these results, showing that β spread by ∼10% but without any noticeable batch-to-batch variation. Having the same characteristics over the two batches assured a negligible ΔV_{BE} variation, which then allowed the use of a low-cost voltage calibration (see Sec. 4.6.4).

The batch-to-batch variation was also evaluated for two batches using UHV_T de-

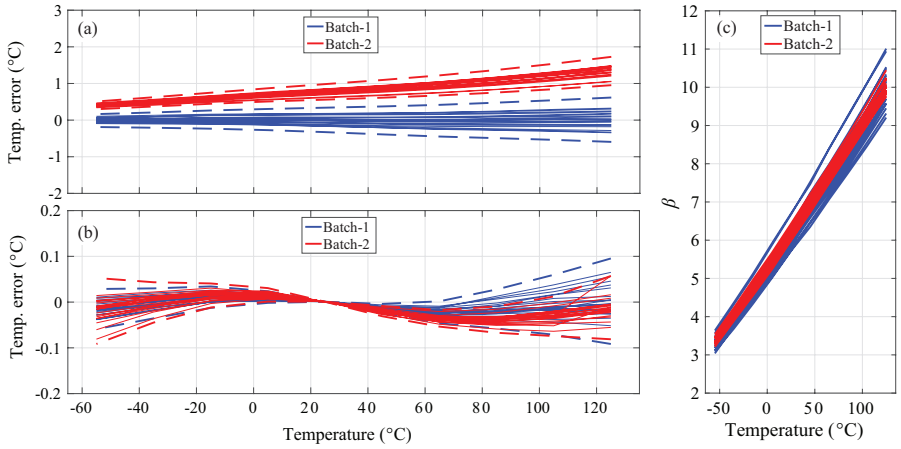


Figure 4.20: Variation in the TDC from batch-to-batch. Batch-1 and Batch-2 use the same calibration parameters extracted from Batch-1: $\alpha = 15.3408$, $A = 615.2447$, $B = -283.0050$, with (a) untrimmed, (b) trimmed and (c) current-gain.

vices (Batch-1 and Batch-3). Similarly, the average parameters were extracted based on Batch-1 and then applied for the two batches. Fig. 4.21 illustrates these measured results and shows that the variation between Batch-1 and Batch-3 was ∼1 °C. A single-point RT trim corrected the batch-to-batch variation and preserved the batch-calibrated inaccuracy: ±0.1 °C (3σ) after the trim. The measured β also shows a negligible batch-to-batch variation, as illustrated in Fig. 4.21 (c).

4.6.3 Switch leakage

Both types of UHV_T and HV_T switches provide sufficiently low leakage for accurate operation up to +125 °C, while normal V_T switches show substantially higher leakage and thus temperature error. To evaluate the effect of switch leakage, the inaccuracy of the three flavors is shown in Fig. 4.22. The UHV_T samples were taken from Batch-1, while HV_T and V_T were taken from Batch-2. All the TDCs used the same average parameters obtained from Batch-1. As shown in Fig. 4.22 (a), at −55 °C the effect of switch leakage is negligible, therefore a 0.5 °C error between the TDCs could be trimmed to less than ±0.1 °C (3σ). Above RT, the effect of switch leakage becomes dominant and causes a 1.4 °C error for the TDCs using normal V_T devices.

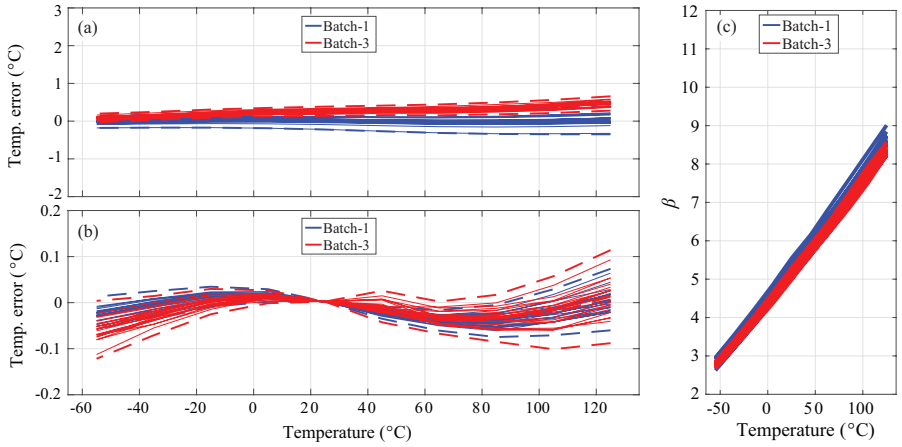


Figure 4.21: Variation in the TDC from batch-to-batch. Batch-1, and Batch-3 use the same calibration parameters extracted from Batch-1: $\alpha = 15.3354$, $A = 618.4909$, $B = -283.6977$, with (a) untrimmed, (b) trimmed and (c) current-gain.

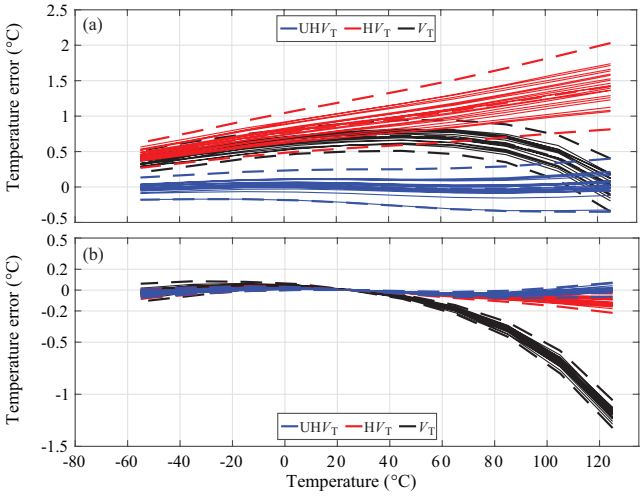


Figure 4.22: Effect of switch leakage on the TDC's accuracy: (a) untrimmed and (b) trimmed.

4.6.4 Voltage calibration (VCAL)

So far, temperature calibration (TCAL) was used to determine X_{OS} for each TDC sample. Temperature calibration is a slow process since TDC samples need to reach thermal equilibrium with the external temperature reference T_{REF}. In practice, this means that the metal block must remain in a thermally isolated chamber for tens

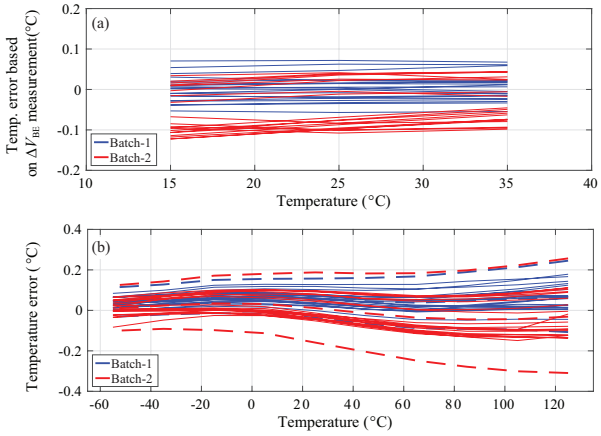


Figure 4.23: Voltage calibration. (a) Die temperature estimation based on the ΔV_{BE} measurements, using the same calibration parameters from Batch-1: $C = 140.27\ \mu V/°C$, $D = -273.73\ \mu V$. (b) Voltage calibrated inaccuracy.

of minutes until the variations in its temperature drop below the desired accuracy. Obtaining ±10 mK error in TCAL requires that both the inaccuracy of T_{REF} and its variation be less than ±10 mK.

Since voltage calibration (VCAL) does not require an accurate T_{REF}, it is a faster alternative to TCAL [4.2],[4.6],[4.7]. In VCAL, the die temperature of each sample T_{die} is determined by comparing its own ΔV_{BE} to an external voltage reference V_{ext}. VCAL relies on the fact that ΔV_{BE}, unlike V_{BE}, is less dependent on process variation, making it an accurate reference.

VCAL involves two steps. In the first step, the TDC digitizes the ratio $X_V = V_{ext}/\Delta V_{BE}$, from which T_{die} can be estimated. This, however, requires the absolute value of V_{ext} and the temperature-dependence of ΔV_{BE} to be known. While V_{ext} is externally determined, the temperature-depence of ΔV_{BE} can be obtained from batch calibration. This involves extracting the parameters C ($\sim 140\mu$V/°C) and D (~ -247 μV), which in a Least-Square fashion, $C \cdot T + D$ fit to the master curve $\Delta V_{BE,avg}(T)$. The estimated die temperature T_V is then found from $T_V = (1/C) \cdot (V_{ext}/X_V - D)$. In the second step of VCAL, the TDC digitizes the ratio $X_{T,untrim}$. Since both steps are completed in quick succession (less than 200 ms), T_{die} can be assumed constant. $X_{T,trim}$ in then found in the same fashion as in (4.7) and (4.8), except that 25 °C is replaced by T_V.

The accuracy of VCAL relies on the accuracy of the voltages V_{ext} and ΔV_{BE}. Errors in V_{ext} directly impact X_V and T_V, thus causing calibration inaccuracy. To

use VCAL to obtain better than ±10 mK accuracy, V_{ext} ($\sim$ 0.65 V) must be accurate to within 0.03%. The accuracy of ΔV_{BE}, can also be enhanced by some circuit techniques. Biasing PNPs at low current levels to suppress R_S-related errors, using DEM to ensure an accurate 1 : 5 emitter current ratio, and chopping Q_L and Q_R to mitigate their mismatch have resulted in T_V inaccuracies better than ±0.15 °C [4.2],[4.24]. The use of VCAL as a low-cost calibration method also requires ΔV_{BE} to be reproducible from one batch to another. Otherwise, C and D need to be obtained as new for each production batch, which then eliminates its advantages.

Fig. 4.23 (a) shows the extracted T_V for the TDCs in Batch-1 and Batch-2 based on the batch-calibrated C and D obtained from Batch-1. To cover a reasonable margin around calibration temperature, T_V is shown in the temperature range from 15 °C to 35 °C. T_V is accurate to less than ±0.15 °C for the samples in Batch-1 and Batch-2. Using the same techniques, an error of less than ±0.1 °C was obtained for the T_V obtained from Batch-1 and Batch-3.

The results of VCAL for the TDCs of Batch-1, and Batch-2 are illustrated in Fig. 4.23 (b), showing a trimmed inaccuracy of ±0.3 °C (3σ). The same inaccuracy was obtained for TDCs of Batch-1 and Batch-3. Although not as accurate as TCAL, VCAL can be implemented at a much lower cost.

4.6.5 Batch trimming

To obtain high accuracy, all samples must first be calibrated (temperature or voltage) and then trimmed. This adds to the testing costs since a wafer might include tens of thousands of TDCs. Alternatively, a batch trim can be done at the expense of extra inaccuracy. In other words, all samples within a batch are trimmed with a single parameter, found by individually trimming a few samples of the batch. The resulting accuracy is then comparable to batch-calibrated accuracy (Fig. 4.18 (a)) but less than that of individual calibration (Fig. 4.18 (b)).

Fig. 4.24 demonstrates the batch trimming that was applied to Batch-1 and Batch-2. The trimming parameter for Batch-2 was a fixed number $X_{OS,B2}$ (= 0.0954) that was found using $X_{OS,B2} = X_{T,avg,B1}(25°C) - X_{T,untrim,B2}(25°C)$, where $X_{T,avg,B1}$ is the master curve of X_T from Batch-1, and $X_{T,untrim,B2}$ is the average of the untrimmed output of several TDCs from Batch-2. This trimming parameter was then added to the ADC outputs in the samples of Batch-2. After the batch-trim, the TDCs in Batch-2 were aligned with those in Batch-1, resulting in an inaccuracy of ±0.4 °C (3σ).

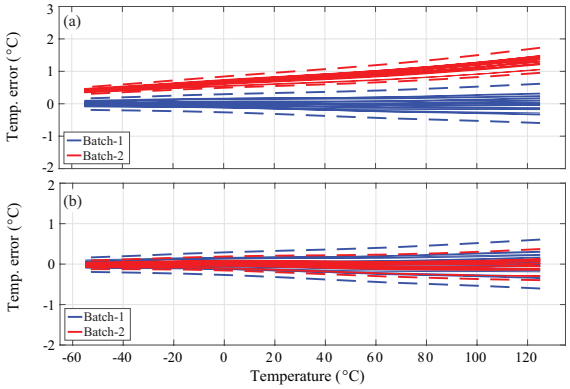


Figure 4.24: Inaccuracy of the TDC using batch trimming: (a) untrimmed and (b) batch-trimmed.

4.7 Conclusion

This chapter described the realization of a TDC in 0.16-μm technology. It obtains a batch-calibrated inaccuracy of ± 60 mK (3σ) from -55 °C to 125 °C after a single-point temperature calibration at room temperature. Unlike previous designs [4.2], the TDC only requires a one-time batch calibration to achieve an inaccuracy of ± 0.1 °C (3σ) from -55 °C to 125 °C, which greatly reduces its calibration costs. To further reduce the costs associated with temperature calibration, alternative approaches such as voltage calibration and batch trimming can be used. These methods reduce calibration costs at the expense of extra inaccuracy. Over multiple batches, voltage-calibration obtained a ± 0.3 °C (3σ) while batch trimming obtained a ± 0.4 °C (3σ) inaccuracy.

Bibliography

[4.1] M. A. P. Pertijs, K. A. A. Makinwa, and J. H. Huijsing, "A CMOS smart temperature sensor with a 3σ-inaccuracy of ± 0.1°C from -55°C to 125°C," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 12, pp. 2805–2815, Dec. 2005.

[4.2] 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°C to 125°C," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 1, pp. 292–301, Jan. 2013.

[4.3] F. Sebastiano, L. J. Breems, K. A. A. Makinwa, S. Drago, D. M. W. Leenaerts, and B. Nauta, "A 1.2-V 10-μW NPN-based temperature sensor in 65-nm CMOS with an inaccuracy of 0.2°C (3σ) from -70°C to 125°C," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 12, pp. 2591–2601, Dec. 2010.

[4.4] A. L. Aita, M. A. P. Pertijs, K. A. A. Makinwa, J. H. Huijsing, and G. C. M. Meijer, "Low-power CMOS smart temperature sensor with a batch-calibrated inaccuracy of ± 0.25°C ($\pm 3\sigma$) from -70°C to 130°C," *IEEE Sensors Journal*, vol. 13, no. 5, pp. 1840–1848, May 2013.

[4.5] G. Wang, A. Heidari, K. A. A. Makinwa, and G. C. M. Meijer, "An accurate BJT-based CMOS temperature sensor with duty-cycle-modulated output," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 2, pp. 1572–1580, Feb. 2017.

[4.6] M. Pertijs and J. Huijsing, "Digital temperature sensors and calibration thereof," Netherlands Patent, Mar., 2010, uS Patent 7,674,035.

[4.7] M. A. P. Pertijs, A. L. Aita, K. A. A. Makinwa, and J. H. Huijsing, "Low-cost calibration techniques for smart temperature sensors," *IEEE Sensors Journal*, vol. 10, no. 6, pp. 1098–1105, June 2010.

[4.8] I. E. Opris, "Series resistance compensation in translinear circuits," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 45, no. 1, pp. 91–94, Jan. 1998.

[4.9] X. Pu, M. Ash, K. Nagaraj, J. Park, S. Vu, P. Kimelman, and S. de la Haye, "An embedded 65 nm CMOS remote temperature sensor with digital beta correction and series resistance cancellation achieving an inaccuracy of 0.4 °C (3σ) from -40 °C to 130 °C," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 9, pp. 2127–2137, Sep. 2015.

[4.10] K. Souri and K. A. A. Makinwa, "A 0.12 mm^2 7.4 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.

[4.11] M. Pertijs and J. Huijsing, "Bias circuits," Netherlands Patent, Nov., 2008, uS Patent 7,446,598 B2.

[4.12] M. A. P. Pertijs and J. H. Huijsing, *Precision temperature sensors in CMOS technology*. Dordrecht, The Netherlands: Springer, Jan. 2006.

[4.13] G. C. M. Meijer, "Thermal sensors based on transistors," *Sensors and Actuators A: Physical*, vol. 10, no. 1-2, pp. 103–125, Sep. 1986.

[4.14] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct 1989.

[4.15] R. P. Fisk and S. M. R. Hasan, "A calibration-free low-cost process-compensated temperature sensor in 130 nm CMOS," *IEEE Sensors Journal*, vol. 11, no. 12, pp. 3316–3329, Dec. 2011.

[4.16] Y. Chae, K. Souri, and K. A. A. Makinwa, "A 6.3 μW 20 bit incremental zoom-ADC with 6 ppm INL and 1 μV offset," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 12, pp. 3019–3027, Dec. 2013.

[4.17] B. Gönen, F. Sebastiano, R. Quan, R. van Veldhoven, and K. A. A. Makinwa, "A dynamic zoom ADC with 109-dB DR for audio applications," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 6, pp. 1542–1550, June 2017.

[4.18] S. Karmakar, B. Gönen, F. Sebastiano, R. van Veldhoven, and K. A. A. Makinwa, "A 280 μW dynamic zoom ADC with 120 dB DR and 118 dB SNDR in 1 kHz BW," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 12, pp. 3497–3507, Dec. 2018.

[4.19] K. Souri and K. A. A. Makinwa, *Energy-efficient smart tempearture sensors in CMOS technology*, 1st ed., ser. Analog Circuits and Signal Processing. Springer International Publishing AG, 2018.

[4.20] R. T. Baird and T. S. Fiez, "Linearity enhancement of multibit $\Delta\Sigma$ A/D and D/A converters using data weighted averaging," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 42, no. 12, pp. 753–762, Dec. 1995.

[4.21] Y. Chae and G. Han, "Low voltage, low power, inverter-based switched-capacitor delta-sigma modulator," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 2, pp. 458–472, Feb. 2009.

[4.22] X. Pu, M. Ash, K. Nagaraj, J. Park, S. Vu, P. Kimelman, and S. De La Haye, "A $\pm$0.4°C accurate high-speed remote junction temperature sensor with digital beta correction and series-resistance cancellation in 65nm CMOS," in *2013 Symposium on VLSI Circuits*, June 2013, pp. C214–C215.

[4.23] C. Lu, S. Ravikumar, A. D. Sali, M. Eberlein, and H. Lee, "An 8b subthreshold hybrid thermal sensor with $\pm$1.07°C inaccuracy and single-element remote-sensing technique in 22nm FinFET," in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, Feb. 2018, pp. 318–320.

[4.24] B. Yousefzadeh, S. Heidary Shalmany, and K. A. A. Makinwa, "A BJT-based temperature-to-digital converter with ±60 mk (3 σ) inaccuracy from −55°C to +125°C in 0.16-μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 4, pp. 1044–1052, Apr. 2017.

Chapter 5

Heater-assisted calibration and trimming of BJT-based TDCs[1]

In this chapter, the effect of packaging stress on BJT-based TDCs will be discussed and experimentally investigated. To do this, the inaccuracy of the high-accuracy TDC described in Chapter 4 was carefully evaluated in both ceramic (low stress) and plastic (high stress) packages. It was observed that the latter causes significant packaging shift, resulting in errors of up to 0.6 °C over the military range. However, by combining batch calibration with an individual room temperature calibration, similar inaccuracy (< 0.25 °C) was obtained with both types of packages.

Compared to calibration at wafer level, however, the need to (batch) calibrate packaged parts significantly increases calibration cost. Since the TDC's characteristics remain quite linear, another alternative is to perform a two-point calibration. To do this, a heater-assisted voltage calibration scheme can be used. This enables low-cost calibration at two temperatures: room temperature and the elevated temperature established by an on-chip heater. This can correct the shifts in the sensor's output caused by the mechanical stress of plastic packaging or to maintain the sensor's accuracy over a wide temperature range. Apart from an accurate off-chip voltage reference, this scheme requires no extra infrastructure, and so significantly reduces calibration cost.

To relax the requirements on the reference, a heater-assisted bandgap trimming scheme was also proposed. This exploits the fact that a flat bandgap voltage V_{BG} can be created by adding a fixed PTAT voltage ($\alpha \cdot \Delta V_{BE}$) to a properly trimmed

[1] B. Yousefzadeh, and K. A. A. Makinwa, "A BJT-based temperature-to-digital converter with ±25 °C 3σ-inaccuracy from −40 °C to +180 °C using heater-assisted calibration voltage calibration," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 2, pp. 369-377, Feb. 2020.

V_{BE}. An on-chip heater then allows the flatness of V_{BG} to be rapidly verified at two temperatures.

5.1 Heater-assisted voltage calibration (HA-VCAL)

5.1.1 Operating principle

Since ΔV_{BE} can be used to accurately estimate T_{die}, voltage calibration (VCAL) can be carried out at any temperature (see Sec. 4.6.4). HA-VCAL exploits this capability by using an on-chip heater (see Fig. 5.1) to rapidly heat the die to an elevated, but not very well-defined, temperature. VCAL can then be carried out at room temperature (T_1) and at the elevated temperature (T_2), resulting in a low-cost two-point calibration.

At the start of HA-VCAL, the heater is off, and $T_{die} = T_1$. The TDC then sequentially outputs T_{1V} and T_{1T}, where T_{1V} is the estimated die temperature obtained from $X_V = V_{ext}/\Delta V_{BE}$, and T_{1T} is the untrimmed output derived from $X_T = V_{BE}/\Delta V_{BE}$. The heater is then turned on, thus elevating T_{die} to T_2, after which the TDC outputs T_{2V} and T_{2T}. The two sets of values (T_{1V}, T_{1T}) and (T_{2V}, T_{2T}) can then be used to trim the TDC.

To avoid significant errors, T_{die} should be stable during VCAL. Due to the thermal time constants of the package, however, the heater-induced step in T_{die} takes several minutes to fully settle, as shown in Fig. 5.2 (a). Rather than waiting for T_{die} to settle, a 1st-order interpolation scheme can be used to mitigate settling-related calibration errors. This involves averaging the value of T_V before (T_{VB}) and after (T_{VA}) a T_T conversion. For consistency, this is done during both phases of HA-VCAL: i.e., during the determination of T_1 and T_2. To evaluate the residual calibration error, the TDC is configured to output T_V continuously during a thermal transient. Interpolated values of T_V [$= 0.5 \cdot (T_{VA} + T_{VB})$] are then compared with their actual values, as shown in Fig. 5.2 (b). Although the interpolation error is initially large, it reaches the noise-level in less than 0.5 s, even for a large change in die temperature (from $\sim$ 25 °C to $\sim$ 90 °C).

5.1.2 On-chip heater

In order to perform HA-VCAL, it is sufficient to heat only the temperature-sensitive elements of the TDC: the four PNPs (Q_R, Q_L, Q_{Rb} and Q_{Lb}) and the two poly-

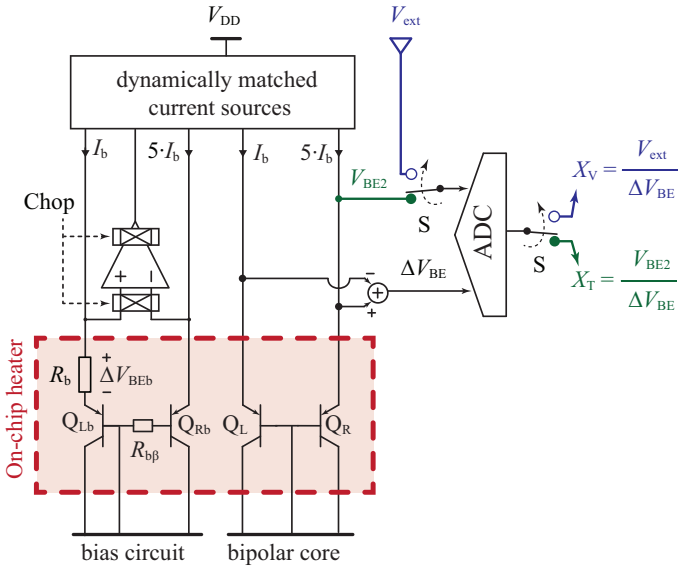


Figure 5.1: Sensing frontend with an on-chip heater. The on-chip heater heats up the temperature-sensitive elements and raises their temperature from T_1 to T_2, while the ADC converts X_T and X_V.

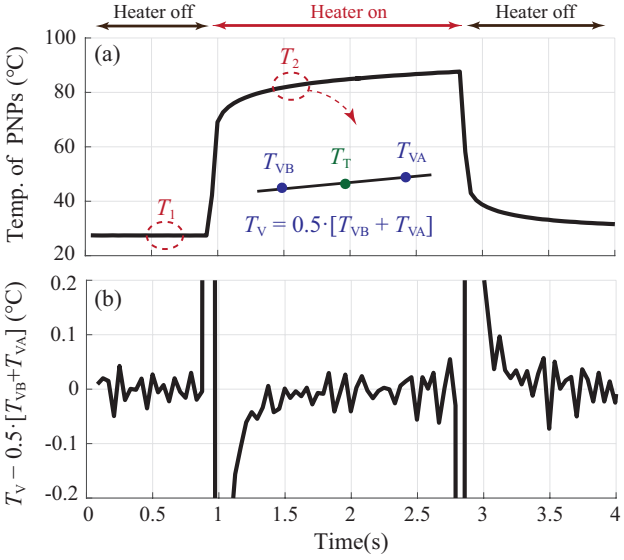


Figure 5.2: Interpolation scheme during a transient heating. (a) Measured temperature of the PNPs. (b) Interpolation error.

resistors (R_b and $R_{b\beta}$) in Fig. 5.1. As a result, the required heater can be quite small, saving both power and area. To facilitate close proximity to the PNPs and the poly resistors, the heater is realized in Metal-2 (Fig. 5.3 (a)), with Metal-1 being reserved for local routing.

On-chip heaters for calibration purposes have been employed in previous work. In [5.1], four on-chip heaters were used to calibrate a frequency reference. To heat the die uniformly, the heaters were placed far away from the circuits being calibrated, and a heat-spreading metal layer was applied to the bottom of the die during a post-fabrication step. In this work, instead of attempting to heat the die uniformly, a single heater is used to create a hot spot directly above the temperature-sensitive circuit elements. As a result, only a small volume of silicon needs to be heated, resulting in faster calibration and obviating the need for an additional heat-spreading layer.

5.1.3 Sources of error

One source of error that is common to both 2-point VCAL and HA-VCAL is the spread of ΔV_{BE}, which results in lower accuracy compared to a conventional 2-point TCAL. In addition, HA-VCAL is less accurate than 2-point VCAL, since there are extra sources of error due to local heating. These errors can originate from temperature-gradients, unheated circuit elements, and leakage currents of sampling switches.

- *Spread of ΔV_{BE}:* In 2-point VCAL, the trimmed output of the TDC, $D_{out,trim}$, is calculated from:

$$D_{out,trim}(T) = A' \cdot \mu_T(T) - B', \tag{5.1}$$

  where $\mu_T(T) = \alpha/(\alpha + X_T(T))$. Assuming that $\mu_T(T)$ is linear, A' and B' are new calibrated parameters that are calculated for each sample based on the two sets of values at T_1 and T_2 as:

$$A' = \frac{T_{2V} - T_{1V}}{\mu_{2T} - \mu_{1T}}, \tag{5.2}$$

$$B' = \frac{T_{1V} \cdot \mu_{2T} - T_{2V} \cdot \mu_{1T}}{\mu_{2T} - \mu_{1T}}, \tag{5.3}$$

  where μ_{1T} and μ_{2T} are the values of μ_T at T_1 and T_2.

  The spread of ΔV_{BE} causes errors in T_{1V} and T_{2V}, which cause errors in A', B', and thus in $D_{out,trim}$. Assuming an error of ΔT_{1V} and ΔT_{2V}, respectively, for T_{1V} and T_{2V}, and taking (5.2), and (5.3) into account, the error in $D_{out,trim}$,

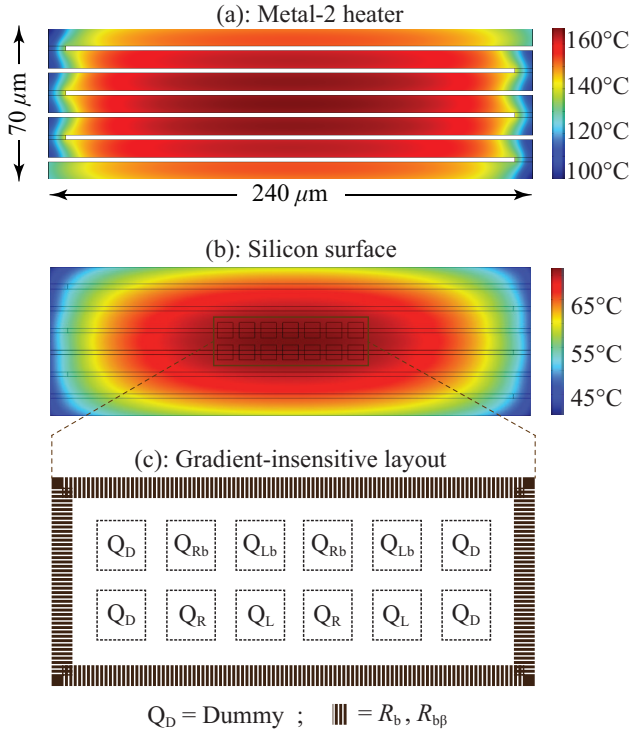


Figure 5.3: Simulated temperature gradient: (a) in the serpentine-shaped Metal-2 heater and (b) on the silicon surface under the heater. (c) Gradient-insensitive layout.

ΔD_{out} can be expressed as:

$$\Delta D_{out}(T) = \frac{\mu_{2T} - \mu_T(T)}{\mu_{2T} - \mu_{1T}} \cdot \Delta T_{1V} + \frac{\mu_T(T) - \mu_{1T}}{\mu_{2T} - \mu_{1T}} \cdot \Delta T_{2V}. \tag{5.4}$$

The key observation from this equation is that the error after trimming will be equal to ΔT_{1V} at T_1, and to ΔT_{2V} at T_2.

Taking into account the spread of T_{1V}, $\sigma_{T_{1V}}$, and the spread of T_{2V}, $\sigma_{T_{2V}}$, and assuming that these two are uncorrelated sources of error, the spread of D_{out}, $\sigma_{D_{out}}$, can be obtained from:

$$\sigma^2_{D_{out}}(T) = \left(\frac{\mu_{2T} - \mu_T(T)}{\mu_{2T} - \mu_{1T}}\right)^2 \cdot \sigma^2_{T_{1V}} + \left(\frac{\mu_T(T) - \mu_{1T}}{\mu_{2T} - \mu_{1T}}\right)^2 \cdot \sigma^2_{T_{2V}}. \tag{5.5}$$

Fig. 5.4 shows the measured result of T_V. Assuming $T_1 = 25°\text{C}$, and $T_2 =$

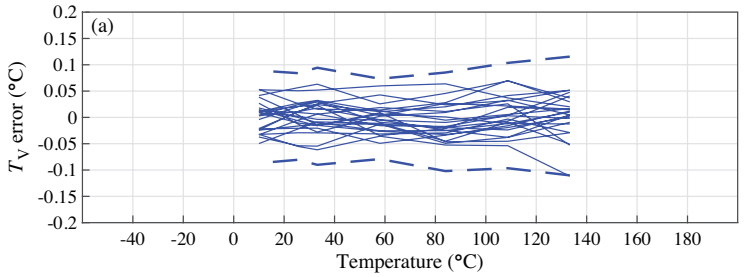


Figure 5.4: Inaccuracy of T_V obtained from the ΔV_{BE} measurment.

110°C, $\pm 3\sigma$ error of $\sim \pm 0.1$°C can be expected for T_V at each of these temperatures. In addition, using nominal values for μ_{T1} (= 0.50), μ_{T2} (= 0.64), and using $\sigma_{T_{1V}}$ and $\sigma_{T_{2V}}$ from Fig. 5.4, $\sigma_{D_{out}}$ can be obtained at other temperatures, which gets larger at the extreme ends of the temperature range. This results in $\pm 3\sigma$ errors of ± 0.21°C at -40°C, and ± 0.17°C at $+180$°C.

- *Temperature gradients:* Using a small on-chip heater causes large temperature gradients in the die. These could cause significant temperature differences between Q_L and Q_R, causing errors in ΔV_{BE}, and thus errors in both T_{2V} and T_{2T}. Similarly, temperature differences between the poly resistors and the PNPs, or between Q_{Lb} and Q_{Rb}, could cause errors in the bias current, and thus errors in V_{BE}, and ultimately errors in T_{2T}. Temperature differences between the PNPs of the bias circuit, and the bipolar core could also cause errors in the β-compensation, again causing errors in V_{BE}, and hence in T_{2T}.

  Fig. 5.3 (a, b) shows the results of a COMSOL simulation of the serpentine-shaped Metal-2 heater while it dissipates 0.5 W. As expected, large temperature gradients can be seen in the heater, as well as in the silicon substrate under the heater. There is also a significant vertical gradient between the heater and the silicon substrate. To mitigate the effect of such gradients, the temperature-sensitive elements of the TDC are placed in a small area (0.002 mm^2) under the much larger heater (0.017 mm^2). In addition, a gradient-insensitive layout is used, as shown in Fig. 5.3 (c). Each PNP is divided into two halves and placed in a common-centroid manner around the heater's symmetry axes. Similarly, all top-level metal routing above the heater is symmetrical with respect to the heater's axes of symmetry.

  To avoid bias current errors, R_b and $R_{b\beta}$ should be at the same temperature as the PNPs. These resistors are therefore placed in a ring around the PNPs, which means that they are somewhat further away from the heater. Being implemented in poly, however, they are actually closer to the Metal-2 heater than the PNPs.

These two effects partially cancel each other and help to reduce the temperature difference between the PNPs and the resistors. Simulations indicate that the average temperature difference is about 7 °C. Since the temperature coefficient of poly resistors is quite low ($\sim$ 65 ppm/°C), this results in a 0.045% error in the value of resistance, which corresponds to a negligible error of $\sim$ 4.5 m°C in T_{2T}.

- *Unheated circuit elements:* In normal operation, the die is at a uniform temperature. During HA-VCAL, however, the die is locally heated, which means that the circuitry directly under the heater is significantly hotter than the rest of the TDC's circuitry, i.e. its front-end and ADC. To prevent calibration errors, the performance of this circuitry must be the same during both modes of operation.

  Temperature-dependent errors in the front-end, such as opamp offset and current ratio mismatch are mitigated by the use of chopping and DEM (Chapter 4). Similarly, ADC errors, such as offset and gain error are mitigated through the use of CDS, system-level chopping, and DEM (Chapter 4). These techniques make both the front-end and the ADC insensitive to temperature gradients. As a result, their performance remains the same during both local and uniform heating.

- *Switch leakage:* The leakage current of the ADC's sampling switches is not mitigated by circuit techniques and increases exponentially with temperature. Compared to normal operation, the temperature gradients created by local heating will result in different leakage current levels, which, as described in Sec. 4.5.2, will change the bias currents of the critical PNPs and cause additional calibration error.

  One way to prevent such errors is to also place the sampling switches under the heater. However, this would increase the required area of the uniformly-heated zone under the heater, and thus require a significantly larger heater. A better solution is to use an ADC architecture that minimizes switch leakage.

5.2 Heater-assisted bandgap trimming (HA-BT)

As discussed in Sec. 4.6.4, achieving ± 10 mK inaccuracy with VCAL requires a $\pm$0.03%-accurate 0.65 V reference voltage. However, such a reference is not readily available in typical test environments.

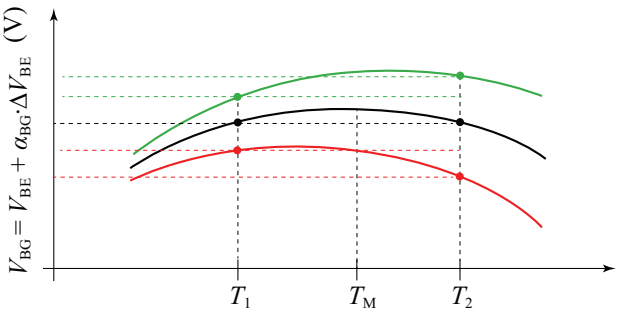


Figure 5.5: Basic principle of bandgap trimming. Nominally $V_{BG}(T_1) = V_{BG}(T_2)$ (shown in black). When I_S spreads, this criterion is no longer valid (shown in green and red).

Heater-assisted bandgap trimming (HA-BT) is a method that trims I_S-related errors without the need to explicitly calibrate V_{BE}, and thus does not require an accurate external reference. In HA-BT, V_{BE} and ΔV_{BE} are combined to create a bandgap voltage, V_{BG}, which is then made flat, i.e., the same at two different die temperatures, by trimming V_{BE}. To minimize the trimming time, RT is chosen as the first temperature, and an on-chip heater is used to rapidly (in less than 0.5 s) generate the second elevated temperature.

In previous work [5.2], [5.3], different ways of using an on-chip heater to perform two-temperature trimming on a bandgap voltage reference have been suggested. They both involve the repeated heating and cooling of the bandgap circuitry and then trimming a number of resistors until a flat bandgap voltage is obtained. In contrast, HA-BT is much faster, because it only involves one heating step, and furthermore, the actual trimming is done in the digital domain.

5.2.1 Operating principle

Bandgap trimming (BT) exploits the fact that if V_{BE} is properly trimmed, a temperature flat bandgap voltage V_{BG} ($= V_{BE} + \alpha_{BG} \cdot \Delta V_{BE}$) can be realized, where α_{BG} is a constant. Taking the curvature of V_{BE} into account, this means that $V_{BG}(T_1)$ ($= V_{1BG}$) will be equal to $V_{BG}(T_2)$ ($= V_{2BG}$), where T_1 and T_2 are properly chosen around the peak value of V_{BG} as shown in the black curve in Fig. 5.5. As shown in this figure, V_{BE} spread causes V_{BG} to spread, such that $V_{1BG} \neq V_{2BG}$. This can be corrected by trimming V_{BE}, resulting in the temperature-flat situation $V_{1BG} = V_{2BG}$.

To implement HA-BT, V_{1BG} and V_{2BG} are digitized by the ADC with respect to

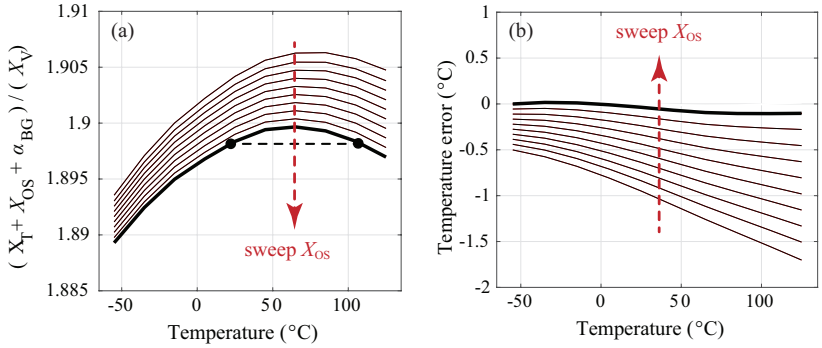


Figure 5.6: Correcting I_S spread using heater-assisted bandgap trimming: (a) $V_{BG,trim}/V_X$ as a result of sweeping X_{OS} and (b) corresponding temperature error after trimming.

an external reference voltage V_X. The trimming criterion then becomes $V_{1BG}/V_X = V_{2BG}/V_X$, which holds for any arbitrary but stable V_X.

As in HA-VCAL, HA-BT uses two pairs of data: (X_{1V}, X_{1T}) at T_1, and (X_{2V}, X_{2T}) at T_2. The corresponding bandgap voltages can then be expressed as:

$$\frac{V_{1BG}}{V_X} = \frac{V_{BE}(T_1) + \alpha_{BG} \cdot \Delta V_{BE}(T_1)}{V_X} = \frac{X_{1T} + \alpha_{BG}}{X_{1V}}\,, \tag{5.6}$$

$$\frac{V_{2BG}}{V_X} = \frac{V_{BE}(T_2) + \alpha_{BG} \cdot \Delta V_{BE}(T_2)}{V_X} = \frac{X_{2T} + \alpha_{BG}}{X_{2V}}\,. \tag{5.7}$$

I_S-related errors in V_{BE} can be trimmed by adding a digital offset X_{OS} to X_T. The correct value of X_{OS} can thus be found by replacing (5.6) and (5.7) in the trimming criterion as in:

$$\frac{X_{1T} + \alpha_{BG} + X_{OS}}{X_{1V}} = \frac{X_{2T} + \alpha_{BG} + X_{OS}}{X_{2V}}\,. \tag{5.8}$$

Solving 5.3 for X_{OS} result in the following expression:

$$X_{OS} = \frac{X_{2T} \cdot X_{1V} - X_{1T} \cdot X_{2V}}{X_{2V} - X_{1V}} - \alpha_{BG}\,. \tag{5.9}$$

The X_{OS} found from (5.9) can then be used to trim the TDC in the same way as shown in (4.7) and (4.9).

Fig. 5.6 illustrates the process of BT using the measured output of a TDC. After choosing $T_1 = 25°C$, and $T_2 = 110°C$, a one-off batch calibration is conducted to determine α_{BG} (~ 14.41) such that the trimming criterion is satisfied for the nominal

value of V_{BE}. It is important to note that $\alpha_{BG} \neq \alpha$, because while α_{BG} establishes a flat bandgap voltage, α makes it slightly PTAT to reduce the non-linearity of μ_T. Fig. 5.6 (a) shows how the digitized bandgap voltage $V_{BG}/V_X = (X_T + X_{OS} + \alpha_{BG})/X_V$ varies over temperature (−55 °C to 125 °C) as it is trimmed with different values of X_{OS}. The corresponding error in $D_{out,trim}$ is shown in Fig. 5.6 (b). It can be seen that the error is minimized, apart from some residual nonlinearity, when the trimming criterion is met. Although Fig. 5.6 uses X_T and X_V over the full temperature range for illustration purposes, only their values at T_1, and T_2 are required to perform HA-BT.

5.2.2 Sources of error

BT is less accurate than TCAL and VCAL since it amplifies the spread of ΔV_{BE}, which causes an additional error. In addition, HA-BT, like HA-VCAL, has to deal with errors associated with local heating. Furthermore, with HA-BT, spread in T_2, e.g. due to the spread of heater resistance will cause additional errors.

- *Spread of ΔV_{BE}:* The spread of ΔV_{BE} causes errors in $X_{1V} = V_X/\Delta V_{BE}(T_1)$ and $X_{2V} = V_X/\Delta V_{BE}(T_2)$, which then causes X_{OS} to deviate from the ideal value found from (5.9). Assuming a spread of ΔX_{1V} and ΔX_{2V}, respectively for X_{1V} and X_{2V}, the resulting X_{OS} can be expressed as:

$$\begin{aligned}\Delta X_{OS} &= \frac{\partial X_{OS}}{\partial X_{1V}} \cdot \Delta X_{1V} + \frac{\partial X_{OS}}{\partial X_{2V}} \cdot \Delta X_{2V} \\ &= \frac{X_{2V} \cdot (X_{2T} - X_{1T})}{(X_{2V} - X_{1V})^2} \cdot \Delta X_{1V} + \frac{X_{1V} \cdot (X_{1T} - X_{2T})}{(X_{2V} - X_{1V})^2} \cdot \Delta X_{2V}\,. \end{aligned} \quad (5.10)$$

  Taking into account the spread of X_{1V}, $\sigma_{X_{1V}}$, and the spread of X_{2V}, $\sigma_{X_{2V}}$, and assuming that they are uncorrelated errors, the spread of X_{OS}, $\sigma_{X_{OS}}$, can be obtained from:

$$\sigma^2_{X_{OS}} = \left(\frac{X_{2V} \cdot (X_{2T} - X_{1T})}{(X_{2V} - X_{1V})^2}\right)^2 \cdot \sigma^2_{X_{1V}} + \left(\frac{X_{1V} \cdot (X_{1T} - X_{2T})}{(X_{2V} - X_{1V})^2}\right)^2 \cdot \sigma^2_{X_{2V}}\,. \quad (5.11)$$

  The spread of $D_{out,trim}$, $\sigma_{D_{out}}$, can then be obtained from:

$$\sigma^2_{D_{out}}(T) = \left(\frac{A \cdot \alpha}{(\alpha + X_T(T))^2}\right)^2 \cdot \sigma^2_{X_{OS}}\,. \quad (5.12)$$

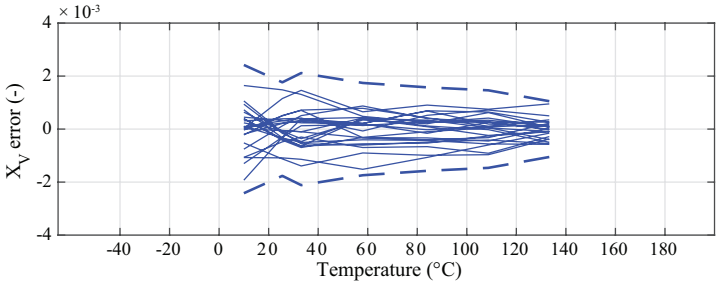


Figure 5.7: Measured spread of X_V for 24 samples from one batch.

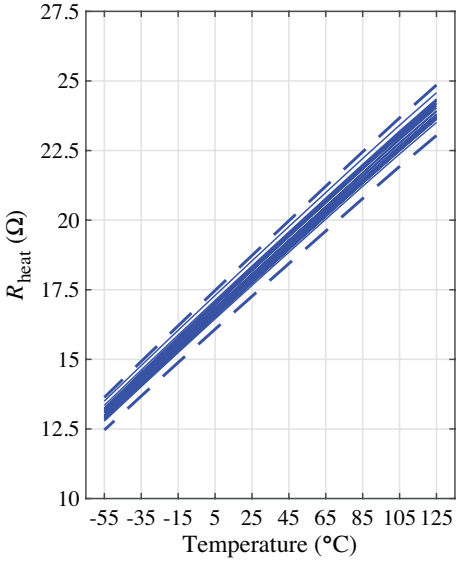


Figure 5.8: Measured resistance of the heater for 24 samples from one batch.

Fig. 5.7 shows the measured results of X_V, which can be used to obtain $\sigma_{X_{1V}}$, and $\sigma_{X_{2V}}$. Assuming the same T_1 and T_2 used in Sec. 5.1.3, and using the nominal values of X_{1T} (= 15.23), X_{2T} (= 8.75), and assuming V_X = 0.3 V to calculate X_{2V} (= 5.65) and X_{1V} (= 7.22), result in worst case $\pm 3\sigma$ errors of ± 0.81°C at +180°C. It is worth noting that according to (5.12), the error is higher at higher temperatures because X_T decreases with temperature [5.4].

- *Spread of heater resistance:* Apart from the spread of ΔV_{BE}, the spread of the heater resistance also causes error in HA-BT. Due to the curvature of V_{BE}, V_{BG} is not perfectly flat over temperature. As a result, variations in the nominal values of T_1 and T_2 will cause trimming errors. In particular, T_2 is a function of the power dissipated by the on-chip heater, P_{heat}, and the thermal resistance of

the package, both of which will spread. The heater's resistance spreads by 10% within a batch (see Fig. 5.8), causing errors of a few degrees in T_2. At T_1, and T_2, however, the slope of V_{BG} is only 38.6 μV/°C, which means that errors of a few degrees in T_1, and T_2 will only cause small (< 1°C) trimming errors.

5.3 Low-leakage readout circuit

Leakage currents in the ADC's sampling network cause errors during heater-assisted calibrations (see Sec. 5.1.3). Reducing these currents should increase both the sensor's accuracy and operating range. A low-leakage TDC was specifically developed to investigate heater-assisted techniques.

5.3.1 Switch leakage

The leakage in the ADC's sampling network (see Fig. 4.15), I_{L1} and I_{L2}, is mainly caused by the leakage currents flowing through the reverse-biased junctions at the source and drain of the switches, and through the off-resistance R_{off} of the other switches connected to C_S. While the former is directly proportional to the total number of switches connected to the PNPs, the latter is a function of $R_{on}/(N \cdot R_{off})$, where N is the number of "off" switches involved.

The zoom ADC used in the TDC described in Chapter 4, has a sampling network consisting of a 64-element capacitor array, and a total of 192 sampling switches. These switches all leak at high temperatures, resulting in a systematic error after HA-VCAL. In other words, the TDC gives slightly different outputs depending on whether the die is heated uniformly to a certain temperature, or whether only the BJTs are heated locally to the same temperature.

In normal operation, leakage currents will also cause errors at high temperatures. To mitigate such errors, BJT-based sensors have either used relatively large bias currents [5.5], or have been realized in low-leakage SOI processes [5.6], [5.7], or have employed leakage compensation schemes [5.8], [5.9].

To suppress the effect of leakage current, a second TDC was designed. It requires only two sampling capacitors and thus much fewer sampling switches. As shown in Fig. 5.9, only 10 switches are required to sample V_{BE1}, V_{BE2} and gnd, as well as to differentially sample V_{ext} via V_{extn} and V_{extp}. Compared to the 192 switches used in the zoom-ADC, this is a substantial reduction. To further reduce the effect of I_{L2}, switch off-resistance is enhanced by using a T-shaped switch configuration

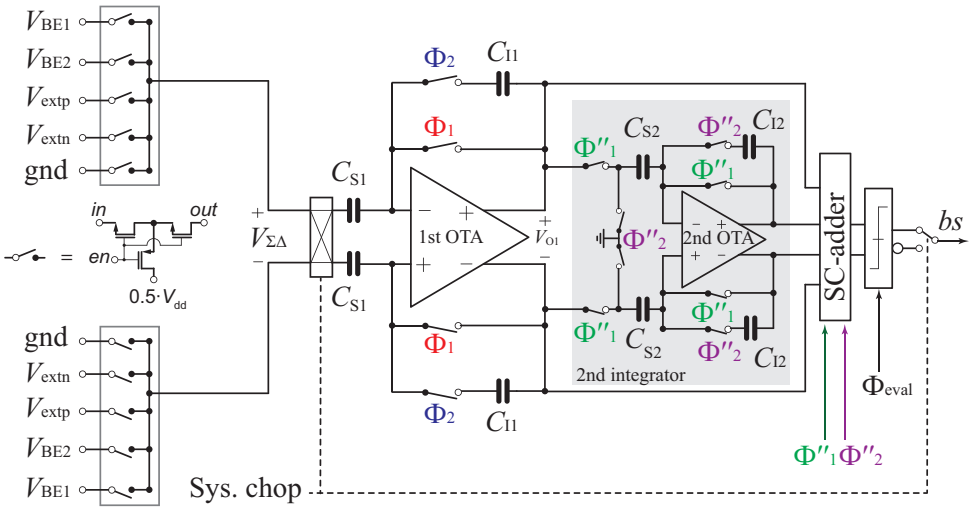


Figure 5.9: Low-leakage readout circuit based on a 2nd-order ΣΔ modulator.

adapted from [5.10]. This consists of two NMOSFETs in series with a PMOSFET that biases the central node to $V_{dd}/2$, when the switch is off (see Fig. 5.9). As a result, one of the two NMOSFETs is always in the deep cut-off region. Compared to TDC_1, the low-leakage ADC used in TDC_2 enables more accurate HA-VCAL at higher temperatures.

5.3.2 Charge-balancing scheme in the low-leakage ADC

Using two sampling capacitors in the low-leakage ADC requires a different charge-balancing scheme than in the zoom-ADC. Similar to [5.11], the ADC digitizes the ratio $Y_T = 3 \cdot \Delta V_{BE}/V_{BE2}$, which varies from ~ 0.15 to ~ 0.72 as T_{die} varies from -40 °C to $+180$ °C. A factor of 3 was chosen to maximize the ADC's dynamic range in this temperature range.

In each cycle of the modulator, the 1st-stage integrates a charge proportional to $3 \cdot \Delta V_{BE}$, when $bs = 0$, or a charge proportional to $3 \cdot \Delta V_{BE} - V_{BE2}$, when $bs = 1$. This results in the desired decimated value: $Y_T = 3 \cdot \Delta V_{BE}/V_{BE2}$. As shown in Fig. 5.10, this charge-balancing scheme is equivalent to setting the ADC's input to $3 \cdot V_{BE2}$ and then balancing it with respect to two temperature-dependent references $3 \cdot V_{BE1}$ and $3 \cdot V_{BE1} + V_{BE2}$.

During VCAL, V_{BE2} is replaced by V_{ext}. The ADC then outputs $Y_V = 3 \cdot$

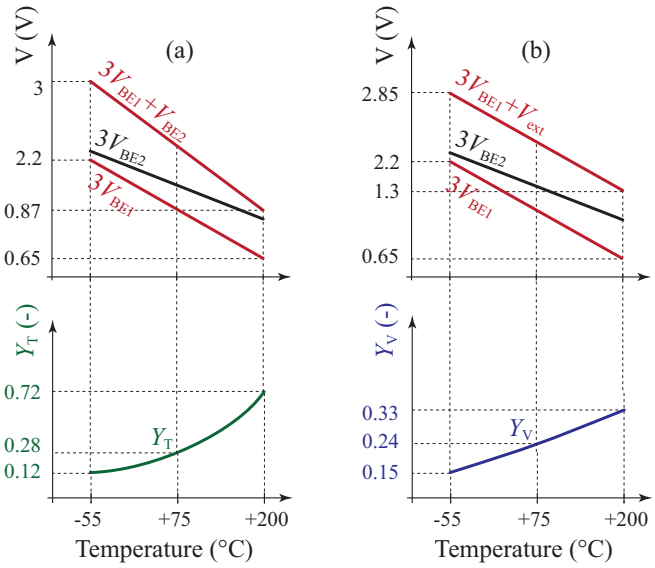


Figure 5.10: Charge-balancing scheme in the low-leakage ΣΔ modulator. The modulator outputs: (a) Y_T in normal mode and (b) Y_V in voltage calibration mode.

$\Delta V_{BE}/V_{ext}$, which varies from ~ 0.15 to ~ 0.33 (for $V_{ext} \sim 0.65$ V) over the operating temperature range. The factor of 3 is maintained for simplicity in the control logic, even though it does not optimize the ADC's dynamic range in this mode.

Compared to the zoom-ADC, the low-leakage ADC requires much simpler control logic, as the ΣΔ conversion is not preceded by a SAR conversion. Also, since the same capacitors C_{S1} sample V_{BE} and ΔV_{BE}, there is no need for DEM, and its corresponding logic. However, the maximum loop-filter input is now $3 \cdot \Delta V_{BE} - V_{BE2}$, which is much larger than the input of $2 \cdot \Delta V_{BE}$ in the zoom-ADC. In order to handle this extra swing with the same current-reuse OTA structure, the first integrator's capacitor was simply increased to 710 fF, i.e. by about 4×.

5.3.3 Sampling scheme in the low-leakage ADC

In order to implement the gain factor (3) required for the charge-balancing scheme, the low-leakage ADC samples either V_{BE}, ΔV_{BE}, or V_{ext} multiple times. As shown in Fig. 5.11, the first integrator uses four non-overlapping clock phases Φ_1, and Φ_2 to sample and transfer the required charge to the integration capacitor C_{I1}. During the fourth phase, the output voltage of the first stage is sampled by the second stage integrator during Φ_1", and then accumulated during Φ_2" to realize a one-cycle

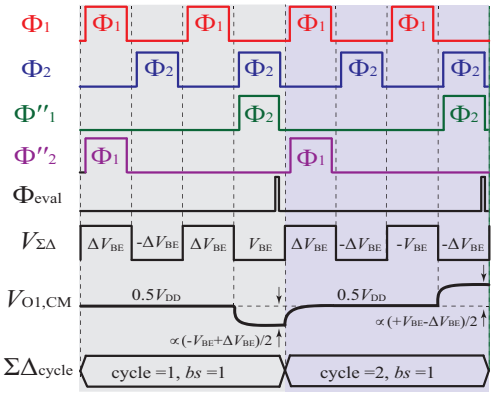


Figure 5.11: Timing diagram of the ΣΔ modulator.

integration delay. The output voltages of the two stages are then summed up by an SC-adder and evaluated by the comparator triggered by Φ_{eval} to generate the output bitstream bs. The sampling capacitor C_{S1} is 125 fF, and each sampling phase takes 1.25 μs, resulting in a 5 μs modulator cycle.

The differential input voltage $V_{\Sigma\Delta}$ of the modulator is shown in Fig. 5.11, for the case when $bs = 1$. The input switches then realize the charge-balancing scheme by applying the following sequence of voltages to $V_{\Sigma\Delta}$: $+\Delta V_{BE}$, $-\Delta V_{BE}$, $+\Delta V_{BE}$, and $+V_{BE2}$. This sequence, however, has a CM component equal to $(V_{BE2} - \Delta V_{BE})/2$, which will also be integrated. Although this component will initially be suppressed by the integrator's CMFB circuit, it is cumulative and may eventually cause the first OTA to clip.

To prevent this, as in [5.12], the CM component is inverted each time the $bs = 1$ state occurs. This involves toggling between the $V_{\Sigma\Delta}$ sequence described above and the following sequence: $+\Delta V_{BE}$, $-\Delta V_{BE}$, $-V_{BE2}$, and $+\Delta V_{BE}$. This results in the same differential voltage, but an inverted CM component, thus driving the integrated CM shift back to zero. Similarly, when $bs = 0$, the $V_{\Sigma\Delta}$ sequence is toggled between: $+\Delta V_{BE}$, $-\Delta V_{BE}$, $+\Delta V_{BE}$, 0; and $+\Delta V_{BE}$, $-\Delta V_{BE}$, 0, $-\Delta V_{BE}$, respectively.

To suppress the residual offset of the first integrator, as well as the mismatch of the two sampling capacitors, the ADC employs system-level chopping. This is implemented by chopping $V_{\Sigma\Delta}$ and digitally inverting the bs polarity. The ADC's final output is then the average of two sub-conversions, each made with different polarity of the system-level chopping signal.

In the low-leakage ADC, the time allocated for V_{BE} sampling is the same as that allocated to ΔV_{BE} sampling. Since ΔV_{BE} is significantly smaller than V_{BE} (about

	Batch	Packaging	Samples	Temp. range	Calibration/Trimming
TDC$_1$	**Batch-1**	Ceramic (DIL-20)	20	−55°C +125°C	HA-VCAL
		Plastic (SO-20)	20		
	Batch-2	Ceramic (DIL-20)	20	−55°C +125°C	HA-VCAL
		Plastic (SO-20)	20		
TDC$_2$	**Batch-3**	Ceramic (DIL-28)	20	−55°C +180°C	HA-VCAL, HA-BT

Table 5.1: Overview of the fabricated TDCs.

16× smaller at 25 °C), this is much longer than is strictly necessary. This, in turn, results in a loss of energy efficiency. As suggested in [5.13], a better approach would be to use different settling times for V_{BE} and ΔV_{BE}, at the expense of a more complex clock generator.

5.4 Experimental results

Two TDCs with on-chip heaters are realized in the 0.16-μm standard CMOS technology. These TDCs use the same frontend and on-chip heater, but different readout circuits. TDC$_1$ uses the zoom-ADC described in Chapter 4, while TDC$_2$ uses the low-leakage ADC described in 5.3. The accuracy of these TDCs are then evaluated using heater-assisted calibration and trimming.

To evaluate the effect of packaging, TDC$_1$ is packaged in both ceramic and plastic. It is also fabricated in two different production batches to observe the effect of packaging variations. HA-VCAL is then used to correct for the shifts observed in the plastic packaging. TDC$_2$ is realized in one production batch and only packaged in ceramic. Compared to TDC$_1$, it achieves a more robust heater-assisted calibration and also operates over a wider temperature range, i.e., up to +180 °C, both of which are mainly due to the reduced leakage currents of its readout circuit. TDC$_2$ is then used to evaluate HA-BT. Table 5.1 shows an overview of the realized TDCs and the corresponding batches, package types, and the calibration and trimming methods used.

TDC$_1$ uses the same zoom-ADC readout circuit described in Chapter 4, however, its layout is altered to accommodate the on-chip heater. Fig. 5.12 (a) shows the chip-micrograph of TDC$_1$, which occupies a total area of 0.17 mm^2, with the on-chip heater taking 0.017 mm^2 of this (i.e., 10%). The low-leakage ADC in TDC$_2$ is smaller than the zoom-ADC and occupies 0.05 mm^2. Fig. 5.12 (b) shows the micrograph of TDC$_2$, which occupies 0.15 mm^2, while accommodating the same on-chip heater.

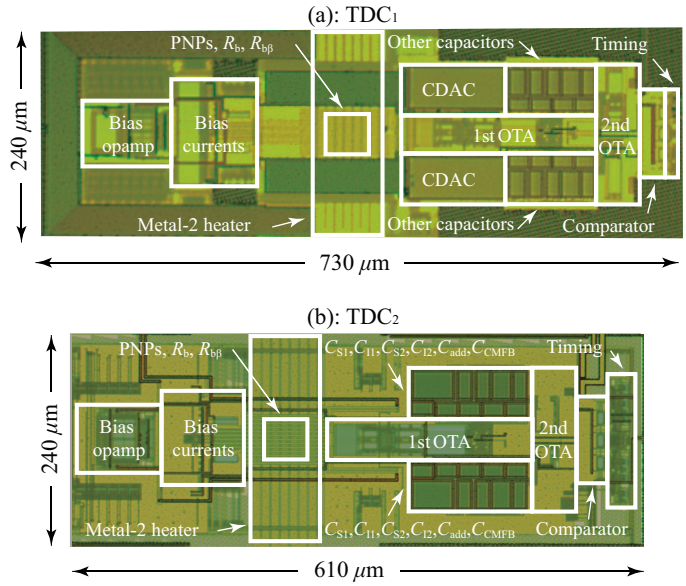


Figure 5.12: Chip micrograph of: (a) TDC_1 and (b) TDC_2.

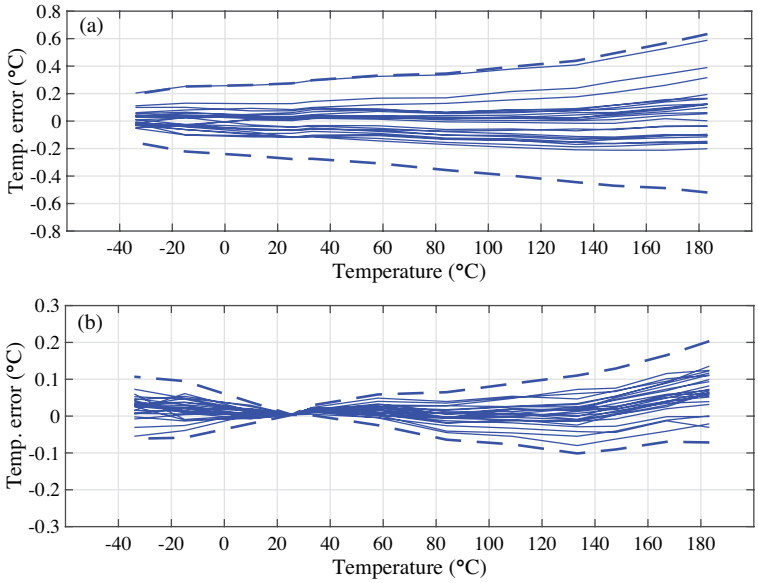


Figure 5.13: Inaccuracy of TDC_2 based on 24 samples from one batch: (a) untrimmed and (b) temperature calibrated and trimmed.

When supplied from a nominal 1.8 V, TDC_2 draws 5.41 μA; the bias currents make up 38%, the bias opamp consumes 19%, and the 1st-stage, the 2nd-stage integrators,

Item	**TDC_2 [5.14]**	**TDC_1 [5.15]**	[5.5]	[5.16]	[5.6]	[5.17]
Sensor type	**BJT**	**BJT**	BJT	BJT	TD [a]	Resistor
Technology (μm)	**CMOS (0.16)**	**CMOS (0.16)**	CMOS (0.16)	CMOS SOI (1)	BiCMOS SOI (0.5)	CMOS (0.18)
T. range (°C)	**−40 to 180**	**−55 to 125**	−55 to 200	25 to 225	−70 to 200	−40 to 180
Inaccuracy (°C)	**±0.2 (3σ)**	**±0.1 (3σ)**	±0.4 (3σ)	±1.6 [b]	±0.4 (3σ)	±0.11 (3σ)
Relative IA (%)	**0.18**	**0.11**	0.31	1.6	0.30	0.10
Calib. (points)	**Temp (1)**	**Temp (1)**	Temp (1)	Temp (1)	Temp (1)	Temp (2)
Power (μW)	**9.75**	**6.9**	35	90	2600	52
Area (mm^2)	**0.15**	**0.17**	0.1	0.41	1	0.12
Res. (mK)	**23**	**7.5**	20	200 [c]	75	0.46
Conv. time (ms)	**20**	**20**	4.2	100	1430	10
FoM (pJK^2) [d]	**103**	**7.8**	59	40×10^3	21×10^6	0.11
Samples	**24**	**20**	16	7	12	20

[a] Thermal diffusivity. [b] Worst case. [c] Based on the reported 10-bit counter.
[d] (Energy per conversion)×(resolution^2).

Table 5.2: Performance summary of the two TDCs and comparison with other high-temperature CMOS TDCs.

and the comparator consume 21%, 14%, and 8%, respectively. For flexibility, the calibration logic, and decimation filter are realized off-chip. TDC_2 achieves an *rms* resolution of 23 mK, in a 20 ms conversion time. This corresponds to a resolution FoM of 103 pJK^2.

To evaluate the accuracy of TDC_2, 24 samples from one batch are packaged in ceramic (DIL-28) and then characterized from −40 °C to +180 °C in a climate chamber. As shown in Fig. 5.13 (a), after the batch calibration and without trimming, the TDC obtains ±0.6 °C (3σ) inaccuracy. Fig. 5.13 (b) shows the error after TCAL and a RT trim, obtaining ±0.2 °C (3σ).

Table 5.2 summarizes the main characteristics of TDC_1 and TDC_2, and compares them with those of a previous zoom-ADC based TDC [5.5], and other high-temperature CMOS TDCs. Although TDC_2 is not as energy-efficient as [5.5], its relative inaccuracy is about 2× better. With the exception of TDC_1, it also consumes significantly less power than other designs.

5.4.1 Errors due to plastic packaging stress

Besides the spread of device parameters, stress-induced errors caused by packaging sets a lower bound on the accuracy of a TDC [5.18]. In Chapter 4, single-point calibration and trimming is used to suppress the process-related errors in V_{BE}. However, the TDC samples used ceramic packaging, which adds a minimum amount of stress to

the TDC dies. Compared to ceramic packaging, plastic packaging cause significantly more die stress which, in turn, causes shifts in V_{BE} after packaging [5.19], [5.20]. To show this, samples of TDC_1 are packaged in both ceramic (DIL-20) and plastic (SO-20) and are then evaluated. 20 samples of the TDC_1 in each package used the same "ceramic" average calibration parameters (i.e., A, B, and α). The individual samples are then calibrated at RT and trimmed. Fig. 5.14 (a) illustrates these results and shows that shifts caused by the plastic packaging led to an error of $\sim$ 0.6 °C after trimming.

As shown in Fig. 5.14 (b), this shift can be reduced to $\pm$0.25 °C (3σ) using new "plastic" average parameters, obtained through a new batch calibration. This involves a multi-temperature calibration of multiple packaged devices. Although a decent accuracy can be obtained, the need to obtain package-specific average parameters followed by a TCAL of the packaged devices is logistically complex and time-consuming, and so significantly increases manufacturing costs. Alternatively, a stress-relieving die coating can be used [5.21], but this would also increase costs.

When the type of plastic packaging and the location of the TDC die in the package are known, the effect of its stress is expected to be reproducible. In this case, the package-specific average parameters can be kept constant and so a one-time batch-calibration is enough to correct for the packaging shift. This is illustrated in Fig. 5.15, which shows the results of TDC_1 in plastic packaging over two different batches (i.e., Batch-1, and Batch-2). Both batches use the same plastic average parameters obtained from Batch-1. Fig. 5.15 (a) shows the untrimmed results, where the TDCs in Batch-2 spreads by $\sim$ 2 °C compared to the ones from Batch-1. This process-related spread can be corrected after post-package individual TCAL and subsequent trimming as illustrated in Fig. 5.15 (b). As a result, the TDCs obtain an inaccuracy of $\pm$0.25 °C (3σ), which is the same as was obtained for Batch-1.

5.4.2 Wafer-level trimming and package-specific calibration

The need to calibrate packaged devices, even at a single point, still adds significantly to the production cost of a TDC, since the time required for packaged devices to reach thermal equilibrium is longer than that of a wafer. In addition, to perform package-level calibration at the same rate as wafer-level calibration, more facilities and infrastructure are required.

When the stress-induced shift of packaging is reproducible, as for the two batches in Fig. 5.15, it is possible to carry out the single-point calibration at wafer level and then apply the package-specific average parameters of the specific package type.

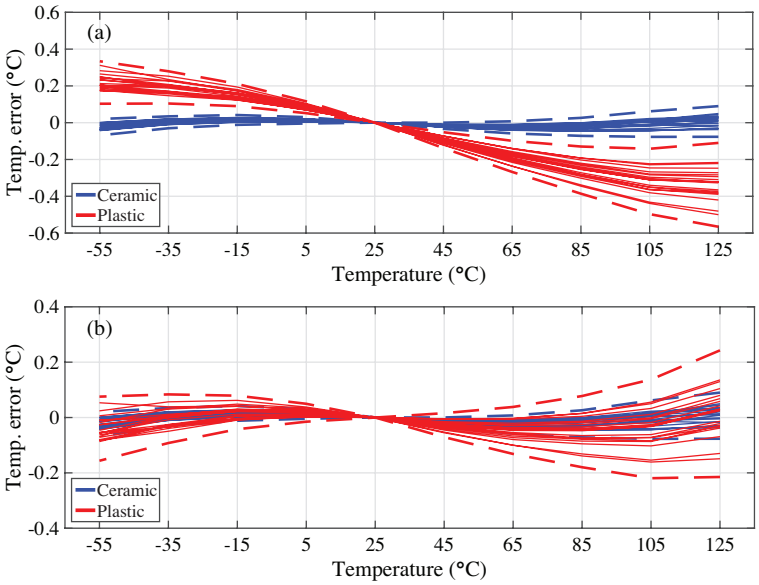


Figure 5.14: Stress-induced shifts caused by plastic packaging on TDC_1: (a) using *ceramic* fitting parameters and (b) using package-specific fitting parameters.

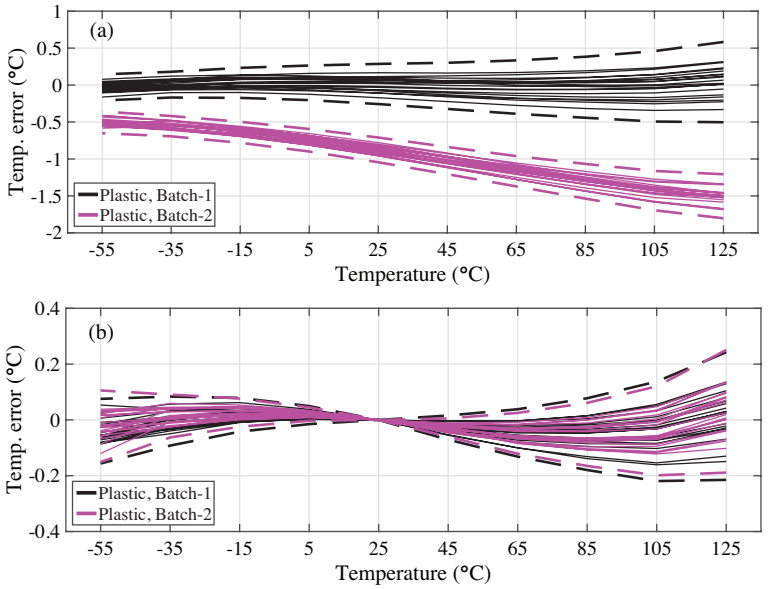


Figure 5.15: Reproducible effect of plastic packaging on TDC_1 over two different batches: (a) untrimmed and (b) temperature calibrated and trimmed.

Demonstrating this requires the availability of some TDC samples for measurement at the wafer level, and then again, the same samples after packaging. This was not an available option. Instead, this is indirectly verified with the help of batch-trimming (Sec. 4.6.5).

Fig. 5.16, illustrates the batch-trimmed results of ceramic packaged devices over the two batches for TDC_1. The parameter that corrects the batch-to-batch spread, X_{OS}, are extracted from these devices. Fig. 5.17 then shows the results of the plastic packaged devices that use the same X_{OS}, but with plastic average parameters. After this, the TDCs obtain ± 0.5 °C (3σ), which are corrected for both spread and stress related errors. In practice, wafer-level trimming can be carried out for individual samples, which then suppresses the spread-related errors more than the batch trimming used in Fig. 5.17.

5.4.3 On-chip temperature gradients

The on-chip heater creates an elevated local temperature, which is then used to calibrate or trim the TDC. To evaluate the actual temperature, the resistance of the heater can be measured as a thermistor. The metal resistance used in the heater had a significant temperature coefficient (0.3%/°C), which makes it a decent temperature sensor. As a result, its average temperature can be accurately determined by measuring its resistance. On-chip Kelvin contacts are used to exclude the series resistances introduced by the bonding wires, PCB traces, and the external switches that drive the heater. As shown in Fig. 5.8, the heater's resistance varies from about $\sim 18\ \Omega$ at 25 °C to about $\sim 23\ \Omega$ at 125 °C, with a $\pm 10\%$ sample-to-sample spread. Calibrating the heater's resistance of a sample leads to its average temperature while calibrating ΔV_{BE} leads to the temperature of the PNPs and then allowed them to be compared during transient heating.

Fig. 5.18 shows the measured results for the temperature of the PNPs and the heater during transient heating when a 5.3 V pulse is applied to the heater. Both the heater and the TDC are clocked by an FPGA, allowing their operation to be accurately synchronized. Initially, the temperature of the heater rapidly rises from 25 °C to 170 °C. The temperature of the PNPs also changes rapidly, but only reaches a maximum of 90 °C, since they are separated from the heater by insulating layers of oxide. The initial rapid rise in temperature is followed by a slow settling phase, which is mainly due to the large thermal time constant of the package.

The temperature difference between the poly resistors and the PNPs is estimated by assuming that the temperature difference ($\sim$ 80 °C) between them is uniformly

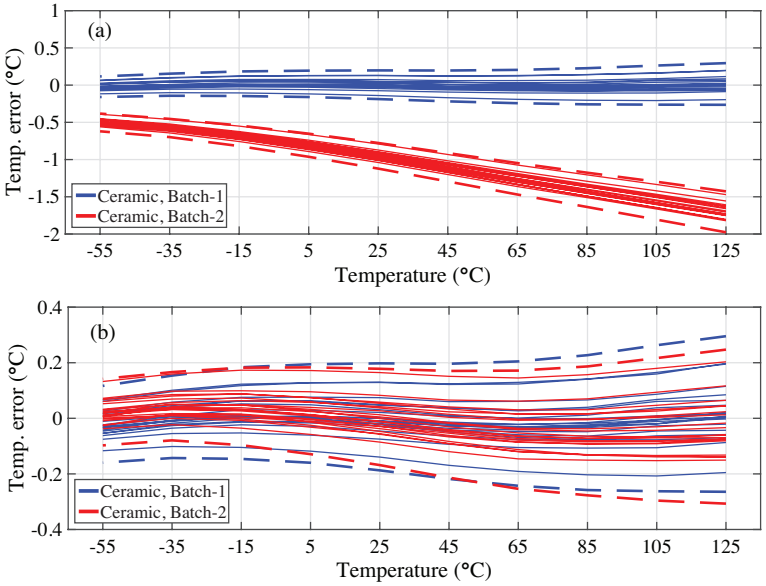


Figure 5.16: Batch trimming of TDC_1 in ceramic packages.

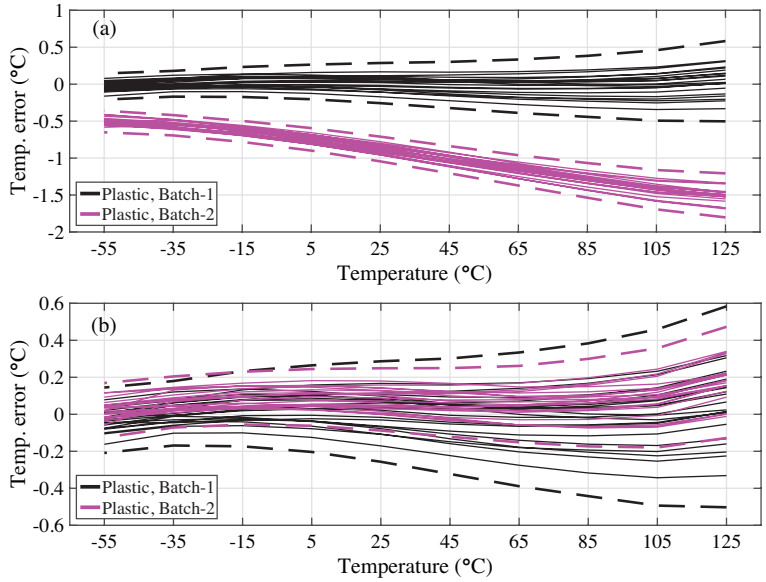


Figure 5.17: Batch trimming of TDC_1 in plastic packaging: (a) untrimmed and (b) batch-trimmed. The TDCs use (1) the trimming parameter, X_{OS}, found from the ceramic-packaged TDCs shown in Fig. 5.16 and (2) *plastic* fitting parameters.

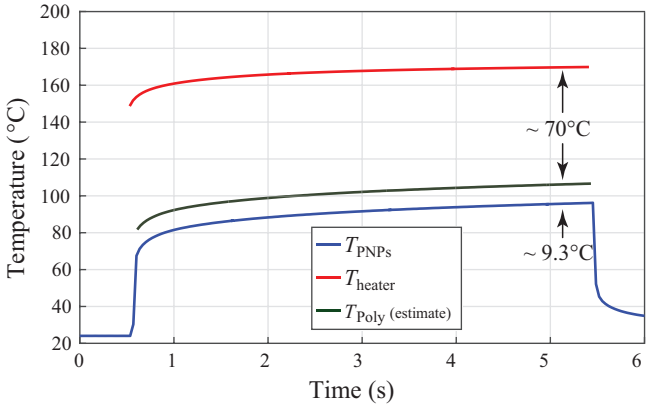


Figure 5.18: Measured temperature of the on-chip heater and the PNPs in transient heating.

distributed over the oxide layers. Based on the layer thicknesses given in the technology datasheet, the estimated temperature difference between the poly resistor and PNPs is 9.3 °C, which correlates reasonably well with the COMSOL simulations (Sec. 5.1.2).

5.4.4 Local heating versus uniform heating

In order to calibrate TDCs with the local heating caused by on-chip heater, the TDC outputs with local heating must be correlated with its output during uniform heating. To evaluate this, first, both temperature calibration and voltage calibration are performed on a TDC while it was uniformly heated in a climate chamber. The resulting outputs, T_T and T_V, respectively, are then used to obtain master curves. These curves are then used to determine the temperature of the TDCs during HA-VCAL. Fig. 5.19 (b) shows that there is an excellent agreement between these two curves (mainly limited by noise of the TDC) in TDC_2. In contrast, in TDC_1, a systematic error of about 1.4 °C was observed as shown in Fig. 5.19 (a). This originates from the large leakage currents in the zoom-ADC. These experiments demonstrate the effectiveness of the low-leakage ADC architecture in mitigating front-end leakage currents.

5.4.5 Conventional VCAL versus HA-VCAL

VCAL, and HA-VCAL are evaluated using 24 samples of TDC_2. While the accuracy of T_V at $T_1 \sim 25$ °C was crucial for VCAL, its accuracy at both $T_1 \sim 25$ °C and T_2

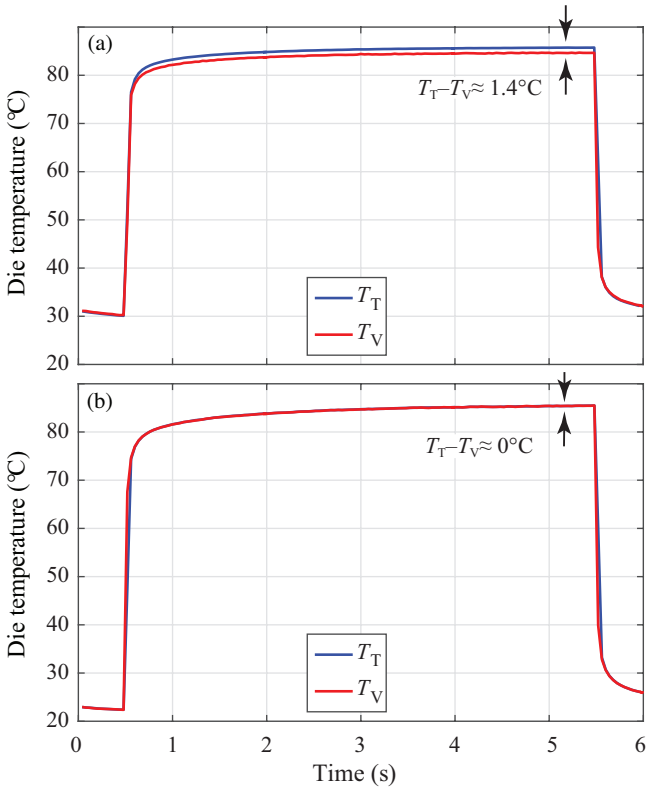


Figure 5.19: Voltage calibrated T_V and temperature calibrated T_T outputs of TDCs during transient heating in: (a) TDC_1 and (b) TDC_2.

$\sim$ 110 °C are important for HA-VCAL. Fig. 5.20 (a) shows the measured T_V from 10 °C to 130 °C. After VCAL at 25 °C, TDC_2 achieved a $\pm$0.3 °C (3σ) inaccuracy from -40 °C to $+180$ °C as shown in Fig. 5.20 (b). This improved to $\pm$0.25 °C using HA-VCAL at $T_1 \sim 25$ °C and $T_2 \sim 110$ °C, as shown in Fig. 5.21 (a). A further increase in T_2, which would be expected to result in higher accuracy, was limited by the heater's maximum current-handling capability.

To further verify the robustness of HA-VCAL, two-point VCAL was performed by uniformly heating the same TDC samples in a climate chamber. Fig. 5.21 (b) shows these results, where the T_1 and T_2 chosen are the same as in Fig. 5.21 (a). The two-point VCAL obtains an inaccuracy of $\pm$0.2 °C (3σ), which is slightly better than the inaccuracy obtained by HA-VCAL. However, this requires significantly more calibration time: tens of minutes versus 0.5 s for HA-VACL.

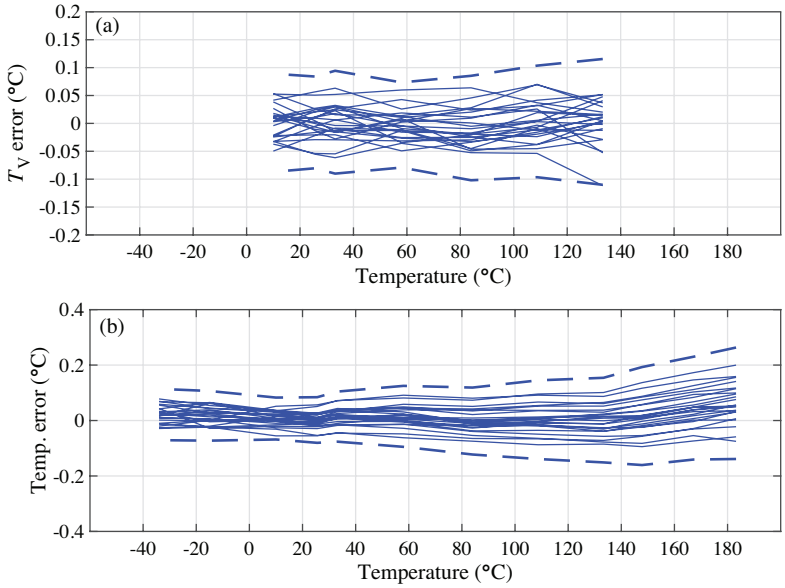


Figure 5.20: Conventional voltage calibration of TDC_2: (a) Inaccuracy of T_V obtained from the ΔV_{BE} measurment and (b) temperature error after voltage calibration at room temperature.

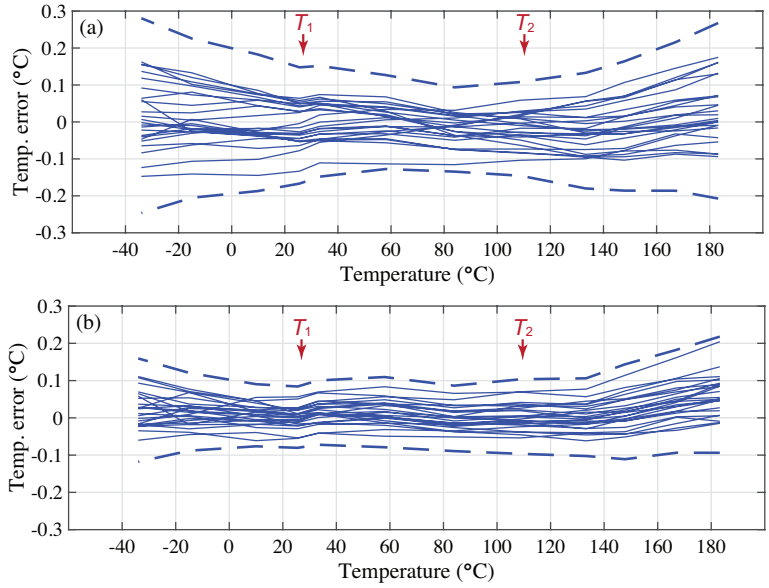


Figure 5.21: Inaccuracy of TDC_2 after using heater-assisted voltage calibration: (a) using the on-chip heater and (b) using a climate chamber.

5.4.6 Post-packaging HA-VCAL

HA-VCAL can also be used to correct for the stress-induced shifts due to packaging. This was evaluated using TDC_1 samples that are packaged in both ceramic and plastic. The packaging shifts depended on the location of the sensing PNPs on the die, as well as on the packaging type and size. For a given package, this could be mitigated with the help of package-specific calibration parameters obtained through batch calibration (Sec. 5.5.3). HA-VCAL, however, provide a low-cost alternative, since it can be applied to any packaging and does not require a temperature-controlled environment.

Measurements on TDC_1 in both ceramic and plastic show that ΔV_{BE} and therefore T_V are reproducible in spite of packaging stress. Fig. 5.22 (a) shows these results, where TDCs in both packages use the same set of average parameters (C, and D) extracted from the ceramic packaged devices to map X_V to T_V.

The TDCs in plastic packaging show a shift of ~ 0.6 °C when they are calibrated at RT and then trimmed (Fig. 5.14 (a)). With the same samples, the results of two-point HA-VCAL at ~ 25 °C and ~ 85 °C are shown in Fig. 5.23 (a). After the calibration and then trimming, the TDC obtains a ± 0.3 °C (3σ) inaccuracy. This is comparable to the results of two-point voltage calibration at 25 °C and 85 °C in a temperature-stabilized climate chamber as shown in Fig. 5.23 (b). It is also comparable to the results of package-specific calibration followed by a RT voltage calibration of the individual samples, as shown in Fig. 5.22 (b). Compared to VCAL alone, HA-VCAL eliminates the need for batch calibration of packaged TDCs.

5.4.7 Heater-assisted bandgap trimming (HA-BT)

The bandgap trimming method was evaluated using 24 samples of TDC_2. After determining A, B, and α through batch calibration, the TDC achieve a ± 0.6 °C (3σ) inaccuracy without trimming, as shown in Fig. 5.24 (a), in blue. This relatively small error is because within the same batch, the spread of I_S is quite limited. Simulations indicate that the worst-case variation may be as large as ± 8 °C (Fig. 4.5).

However, TDC_2 was only available in one production batch. To emulate a large change in I_S, the bias current used to generate V_{BE2} was deliberately changed. With reference to Fig. 5.1, this involves using V_{BE1} (biased at I_b), instead of V_{BE2} (biased at $5 \cdot I_b$) to compute X_T, which is equivalent to a $5\times$ increase in I_S. This drastic change causes a significant error in the TDC's output as shown in Fig. 5.24 (a), in red. As shown in Fig. 5.24 (b), however, this can be reduced to ± 0.2 °C (3σ) after TCAL and then RT trimming. These results show good agreement with the results

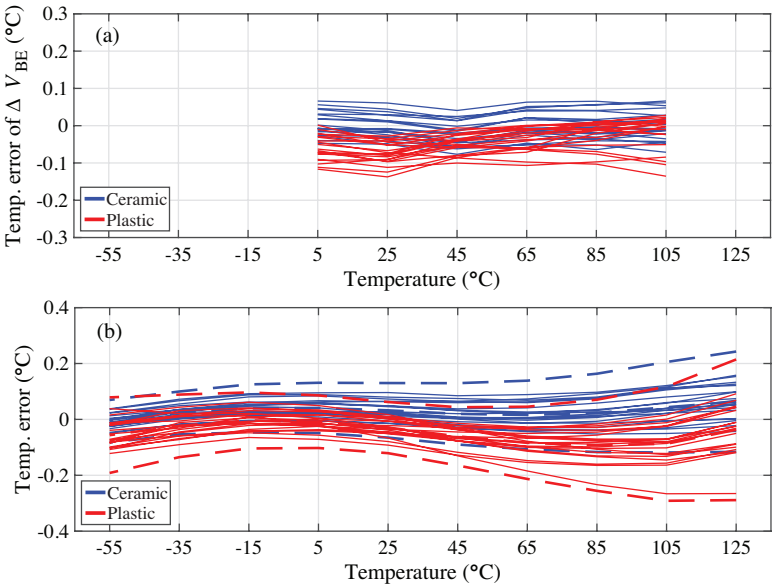


Figure 5.22: VCAL of TDC_1 in ceramic and plastic packaging. (a) The T_V error is based on the ΔV_{BE} measurements using ceramic fitting parameters. (b) The temperature error after VCAL at RT using packaging-specific fitting parameters.

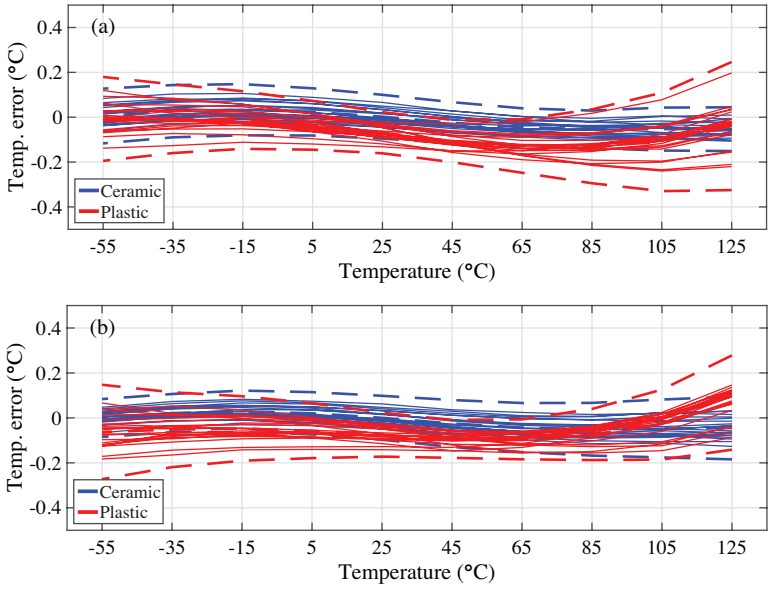


Figure 5.23: Use of two-point heater-assisted voltage calibration to correct packaging shift: (a) using the on-chip heater and (b) using a climate chamber.

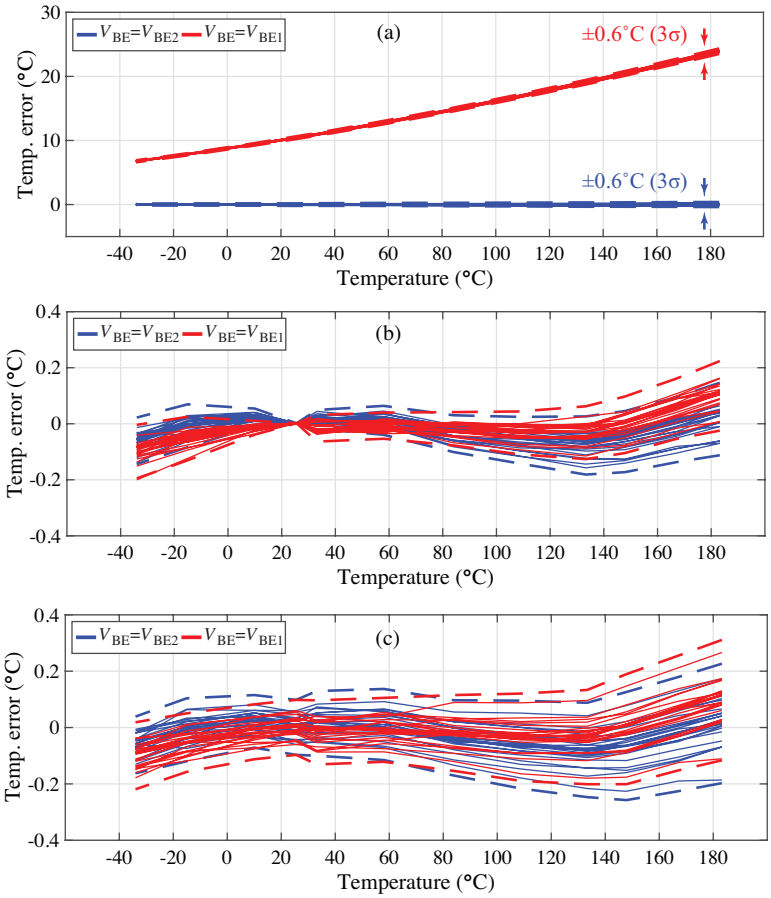


Figure 5.24: Inaccuracy of TDC_2 with a deliberate shift in I_S to emulate batch-to-batch variation: (a) untrimmed, (b) temperature calibrated at room temperature and (c) voltage calibrated at room temeprature.

of the different batches shown in Fig. 4.20 and Fig. 4.21.

VCAL (at $\sim$ 25 °C) was also carried out in the two cases ($V_{BE} = V_{BE2}$ and $V_{BE} = V_{BE1}$). As shown in Fig. 5.24 (c), the TDC's inaccuracy is restored to ±0.3 °C (3σ) after trimming. A small error in V_X results in a large error in the sensor's trimmed output. This is shown in Fig. 5.26 (a), where the measured temperature error of one sample after VCAL at ($\sim$ 25 °C) is shown for different values of V_X. A ±1% error in V_X causes a ±7 °C temperature error. This corresponds to ±7 °C/% to V_X, which is measured to be at a similar level for HA-VCAL.

HA-BT was carried out for the same two cases, the results of which are shown in Fig. 5.25 (a). T_1 is at $\sim$ 25 °C and $T_2 \sim$ 110 °C, which is accomplished by driving the

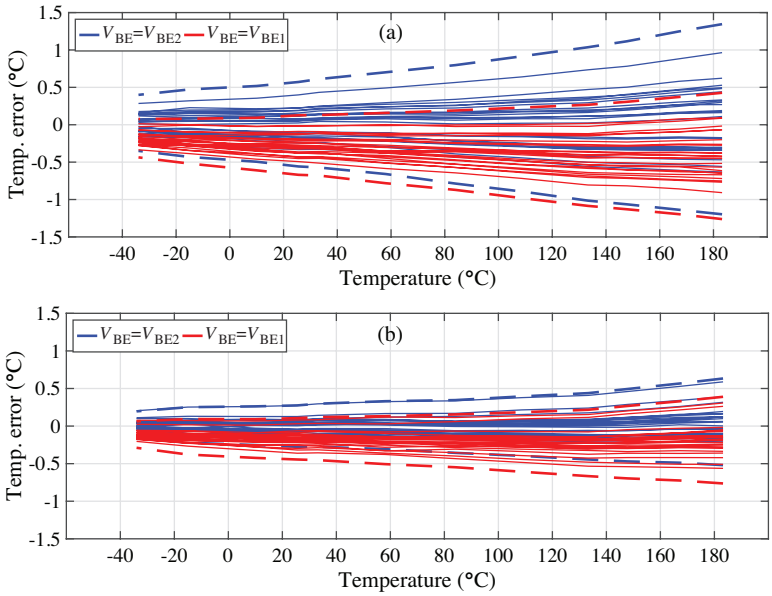


Figure 5.25: Heater-assisted bandgap trimming of TDC_2 using: (a) the on-chip heater and (b) a climate chamber.

on-chip heater with a 6.5 V external voltage. In both cases, the sensor's inaccuracy is restored to less than ±1.3 °C (3σ) after HA-BT. To investigate the effect of heater spread, BT is also done in a climate chamber, which ensures that T_2 is well-defined. As shown in Fig. 5.25 (b), the resulting inaccuracy is ±0.8 °C (3σ), which is better than that obtained with HA-BT. However, this is obtained at the expense of more trimming time than the 0.5 s required for HA-BT. Since the same V_X is used for both BT and HA-BT, the extra ±0.5 °C (3σ) error is associated with the spread of the heater's resistance and T_2.

The main advantage of HA-BT is its lower sensitivity to external values (V_X and P_{heat}) compared to other calibration and trimming methods, which greatly relaxes its requirements. Fig. 5.26 (b) shows the measured sensitivity of HA-BT to V_X. A ±5% error is added to the value of V_X, which then results in a temperature error of less than ±0.5 °C. This corresponds to a sensitivity of 0.1 °C/%, which is 70× lower than that of VCAL. Fig. 5.26 (c) illustrates the measured sensitivity of HA-BT to P_{heat}. In this figure, P_{heat} is varied by ±5% by changing the driving voltage of the heater. This, in turn, causes temperature errors of less than 0.75 °C, corresponding to a sensitivity of 0.075 °C/%. The sensitivity to P_{heat} is, thus, lower than the sensitivity to V_X.

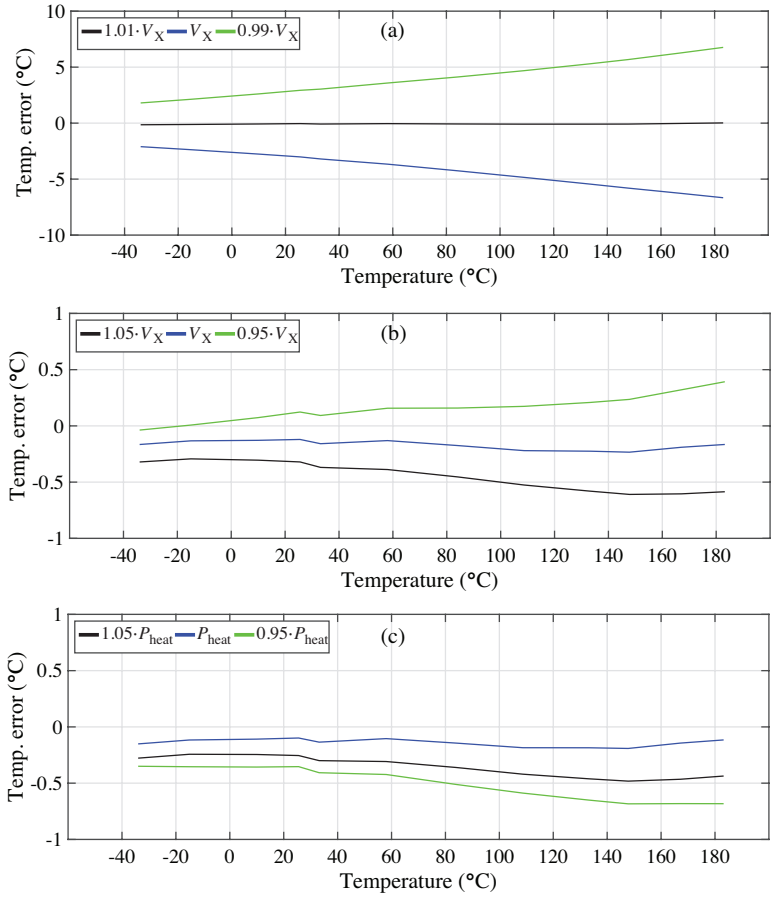


Figure 5.26: Measured sensitivity to external values: (a) Sensitivity of voltage calibration to V_X. Sensitivity of heater-assisted bandgap trimming to: (b) V_X and (c) P_{heat}.

5.5 Conclusion

The effect of mechanical stress caused by plastic packaging on BJT-based TDCs has been studied in this chapter. Compared to ceramic packaging, plastic packaging causes temperature errors of $\sim \pm 0.6$ °C over the military range. These errors can be reduced to $\sim \pm 0.25$ °C by using package-specific average parameters obtained by batch calibration of the packaged devices, when the package size, and type, and the location of the TDC in the package remain the same. Heater-assisted voltage calibration (HA-VCAL) has been introduced as a low-cost alternative, which can be applied to any package. It provides a low-cost two-point calibration by using an on-

chip heater that elevates die temperature from ~ 25 °C to ~ 110 °C within 0.5 s with the help of a $\pm 0.03\%$-accurate external voltage reference. This method can reduce the stress-induced errors of plastic packaging to $\sim \pm 0.3$ °C.

To carry out a robust HA-VCAL, the TDC's circuitry should be insensitive to the local-heating created by the heater. This means that the temperature-dependent errors associated with the TDC's readout circuitry must be suppressed. While many circuit-related errors can be reduced by dynamic techniques, this is not the case for leakage-induced errors. To address this, a TDC with a low-leakage ADC has been introduced, which reduces the number of critical sampling switches from 192 in a TDC based on the zoom-ADC to 10. This low-leakage TDC resulted in a robust operation of the HA-VCAL and operates up to +180 °C with nearly 2× better inaccuracy.

Finally, a new bandgap trimming method is introduced, which unlike the previous calibration method does not require accurate external references for correcting the process-induced spread of V_{BE}. It trims V_{BE} until a flat bandgap voltage is achieved, i.e., one that is the same at room temperature and at an elevated temperature. An on-chip heater is then used to realize a heater-assisted bandgap trimming scheme. Although the resulting inaccuracy (± 1.3 °C) is greater than that obtained with conventional temperature calibration (± 0.2 °C) or voltage calibration (± 0.3 °C), its relaxed requirement on the external reference makes it suitable for noisy test environments. Compared to voltage calibration, it is 70× less sensitive to the absolute value of the external reference.

Bibliography

[5.1] Y. Satoh, H. Kobayashi, T. Miyaba, and S. Kousai, "A 2.9mW, +/- 85ppm accuracy reference clock generator based on RC oscillator with on-chip temperature calibration," in *2014 Symposium on VLSI Circuits Digest of Technical Papers*, June 2014, pp. 1–2.

[5.2] G. Ceremonesi and S. Rossi, "Electrical system, voltage reference generation circuit, and calibration method of the circuit," Patent, Oct. 25, 2011, uS Patent 8044 677 B2.

[5.3] Kee-Chee Tiew, J. Cusey, and R. Geiger, "Inflection point correction for voltage references," in *Proceedings of the 2003 International Symposium on Circuits and Systems, 2003. ISCAS '03.*, vol. 1, May 2003, pp. 649–652.

[5.4] 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°C to 125°C," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 1, pp. 292–301, Jan 2013.

[5.5] K. Souri, K. Souri, and K. Makinwa, "A 40μW CMOS temperature sensor with an inaccuracy of ± 0.4°C (3σ) from -55°C to 200°C," in *2013 Proceedings of the ESSCIRC (ESSCIRC)*, Sep. 2013, pp. 221–224.

[5.6] C. van Vroonhoven, D. D'Aquino, and K. Makinwa, "A ± 0.4°C (3σ) -70°C to 200°C time-domain temperature sensor based on heat diffusion in Si and SiO2," in *2012 IEEE International Solid-State Circuits Conference*, Feb. 2012, pp. 204–206.

[5.7] B. Rue and D. Flandre, "A SOI CMOS smart high-temperature sensor," in *2007 IEEE International SOI Conference*, Oct. 2007, pp. 111–112.

[5.8] K. Mizuno, N. Ohta, F. Kitagawa, and H. Nagase, "Analog CMOS integrated circuits for high-temperature operation with leakage current compensation," in *1998 Fourth International High Temperature Electronics Conference. HITEC (Cat. No.98EX145)*, June 1998, pp. 41–44.

[5.9] J. Nilsson, J. Borg, and J. Johansson, "Leakage current compensation for a 450 nW, high-temperature, bandgap temperature sensor," in *2015 22nd International Conference Mixed Design of Integrated Circuits Systems (MIXDES)*, June 2015, pp. 343–347.

[5.10] K. Ishida, K. Kanda, A. Tamtrakarn, H. Kawaguchi, and T. Sakurai, "Managing subthreshold leakage in charge-based analog circuits with low-V_{TH} transistors by analog T-switch (AT-switch) and super cut-off CMOS (SCCMOS)," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 4, pp. 859–867, April 2006.

[5.11] Z. Tang, Y. Fang, X. Yu, Z. Shi, and N. Tan, "A CMOS temperature sensor with versatile readout scheme and high accuracy for multi-sensor systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 11, pp. 3821–3829, Nov. 2018.

[5.12] F. Sebastiano, L. J. Breems, K. A. A. Makinwa, S. Drago, D. M. W. Leenaerts, and B. Nauta, "A 1.2-V 10-μW NPN-based temperature sensor in 65-nm CMOS with an inaccuracy of 0.2°C (3σ) from -70°C to 125°C," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 12, pp. 2591–2601, Dec. 2010.

[5.13] S. Zaliasl, J. C. Salvia, G. C. Hill, L. Chen, K. Joo, R. Palwai, N. Arumugam, M. Phadke, S. Mukherjee, H. Lee, C. Grosjean, P. M. Hagelin, S. Pamarti, T. S. Fiez, K. A. A. Makinwa, A. Partridge, and V. Menon, "A 3 ppm 1.5 $\times$ 0.8 mm^2 1.0 μA 32.768 kHz MEMS-based oscillator," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 1, pp. 291–302, Jan. 2015.

[5.14] B. Yousefzadeh and K. A. A. Makinwa, "A BJT-based temperature-to-digital converter with a $\pm$0.25°C 3σ-inaccuracy from -40°C to $+180$°C using heater-assisted voltage calibration," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 2, pp. 369–377, Feb 2020.

[5.15] B. Yousefzadeh and K. A. A. Makinwa, "A BJT-based temperature sensor with a packaging-robust inaccuracy of $\pm$0.3°C (3σ) from -55°C to 125°C after heater-assisted voltage calibration," in *IEEE International Solid-State Circuits Conference (ISSCC)*, Feb. 2017.

[5.16] J. Pathrose, C. Liu, K. T. C. Chai, and Y. Ping Xu, "A time-domain band-gap temperature sensor in SOI CMOS for high-temperature applications," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 62, no. 5, pp. 436–440, May 2015.

[5.17] S. Pan, Ç. Gürleyük, M. F. Pimenta, and K. A. A. Makinwa, "A 0.12 mm^2 wien-bridge temperature sensor with 0.1°C (3σ) inaccuracy from -40°C to 180°C," in *IEEE International Solid- State Circuits Conference - (ISSCC)*, Feb. 2019, pp. 184–186.

[5.18] B. Abesingha, G. A. Rincon-Mora, and D. Briggs, "Voltage shift in plastic-packaged bandgap references," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 49, no. 10, pp. 681–685, Oct. 2002.

[5.19] H. Ali, "Stress-induced parametric shift in plastic packaged devices," *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part B*, vol. 20, no. 4, pp. 458–462, Nov. 1997.

[5.20] Y. Jiang and J. Ju, "The influence of plastic-package on the voltage shift of voltage reference in analog circuit," in *2009 International Conference on Electronic Packaging Technology High Density Packaging*, Aug. 2009, pp. 1233–1235.

[5.21] M. Motz, U. Ausserlechner, and M. Holliber, "Compensation of mechanical stress-induced drift of bandgap references with on-chip stress sensor," *IEEE Sensors Journal*, vol. 15, no. 9, pp. 5115–5121, Sep. 2015.

Chapter 6

A Multi-function sensor with temperature, voltage, and capacitance sensing modes [1]

Calibration and trimming costs make up a significant part of the total production cost of a temperature-to-digital converter (TDC). While chapters 4 and 5 focused on ways of reducing these costs, the focus of this chapter is on adding more functionality to a TDC, thus making more cost-effective use of its area.

The area of BJT-based TDCs varies widely, ranging from 0.2 mm^2 for high accuracy sensors, such as the one described in chapter 4, to less than 0.005 mm^2 for low accuracy sensors intended for thermal management [6.1]–[6.3]. In this chapter, rather than compromising accuracy by reducing area, voltage, and capacitive sensing functionality are added to an existing TDC. This is an effective approach for reducing area-related costs, for instance, in IoT sensor nodes [6.4], which are often required to sense multiple parameters sequentially.

6.1 Overview of the multi-function sensor

Fig. 6.1 shows a simplified block diagram of the proposed multi-function sensor. Its main building blocks are (1) a zoom-ADC that is reused for all three sensing modes, (2) a sensing front-end that generates V_{BE} and ΔV_{BE} for the voltage and temperature sensing modes and (3) some multiplexers (S_1:S_4) that are configured according to the mode of operation.

[1] B. Yousefzadeh, W. Wu, B. Butter, K. A. A. Makinwa, and M. Pertijs "A compact sensor readout circuit with combined temperature, capacitance and voltage sensing functionality," *in 2017 Symposium on VLSI Circuits*, pp. C78-C79, June 2016.

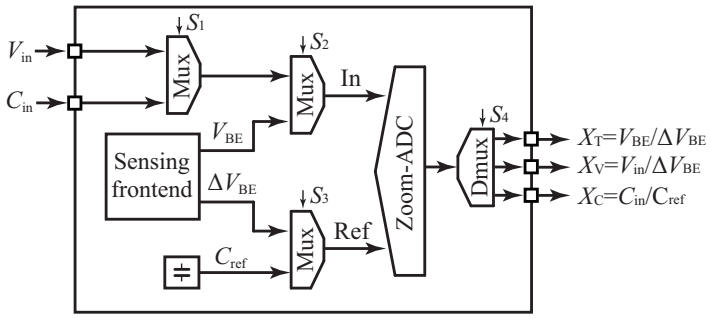


Figure 6.1: Simplified block diagram of the multi-function sensor. It uses a zoom-ADC to convert X_T in temperature sensing, X_V in voltage sensing and X_C in capacitance sensing modes.

In temperature sensing mode, the zoom-ADC digitizes $X_T = V_{BE}/\Delta V_{BE}$ using voltages derived from a BJT-based front-end. As discussed in Chapter 4, X_T can then be mapped to absolute temperature. In voltage sensing mode, the readout circuit digitizes $X_V = V_{in}/\Delta V_{BE}$, in which V_{in} is an input voltage and ΔV_{BE} is a reference voltage. Since ΔV_{BE} is a linear function of temperature and is also quite stable over different batches and packaging, its absolute value can be estimated when the temperature is known. Using the estimated ΔV_{BE}, V_{in} can then be found. Voltage sensing mode thus relies on the fact that the temperature variation between the digitization of X_T and X_V is negligible. When temperature varies at high rates, it is possible to adapt the interpolation scheme described in Section 5.1, to digitize a sequence of X_{TB}, X_V, X_{TA}, and then use the average of X_{TB} and X_{TA} as a measure of temperature.

In capacitance sensing mode, the readout circuit digitizes $X_C = C_{in}/C_{ref}$, in which C_{in} is an input capacitance and C_{ref} is an internal reference capacitance. The latter is the sampling capacitor, C_S, of the zoom-ADC. Realized as a Metal-Insulator-Metal (MIM) fringe capacitor, C_S is quite stable ($\sim \pm 15$ ppm/°C). Its absolute value can thus be found with a single point calibration at room temperature. C_{in} can then be found using X_C and the calibrated value of C_S.

6.2 Temperature-to-digital conversion

For temperature sensing, the readout circuit operates in the same way as the TDC described in Chapter 4. The sensing front-end generates temperature-dependent voltages V_{BE} and ΔV_{BE}, while the zoom-ADC digitizes the ratio X_T ($= V_{BE}/\Delta V_{BE}$).

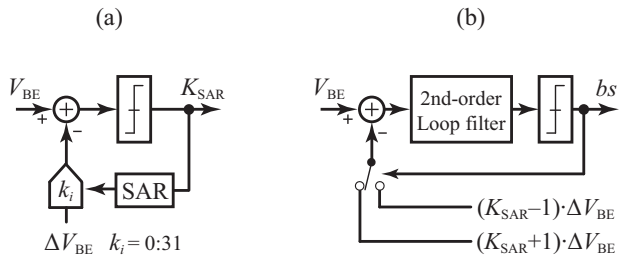


Figure 6.2: Two-step conversion in temperature sensing mode: (a) SAR finds K_{SAR} as the closest integer to X_T and (b) the $\Sigma\Delta$ modulator finds $\mu_{\Sigma\Delta}$.

As shown in Fig. 6.2, this is accomplished in a two-step SAR-$\Sigma\Delta$ conversion. In the first step, the SAR finds the closest possible integer to X_T (K_{SAR}). In the second step, a $\Sigma\Delta$ modulator uses K_{SAR} and balances V_{BE} against reference signals $(K_{SAR}-1)\cdot\Delta V_{BE}$ and $(K_{SAR}+1)\cdot\Delta V_{BE}$. X_T can then be obtained using K_{SAR} and the decimated output of the $\Sigma\Delta$ ($\mu_{\Sigma\Delta}$), using $X_T = K_{SAR} + 2\cdot\mu_{\Sigma\Delta}$.

The SAR conversion takes six clock cycles with non-overlapping phases Φ_1 and Φ_2 to find K_{SAR}. In the first five cycles, during Φ_1, C_{I1} is reset and a charge equal to $C_S\cdot K_i\cdot(-\Delta V_{BE})$ and $C_S\cdot 1\cdot V_{BE}$ is sampled on the CDAC capacitors at differential branches, where K_i $(= 1:31)$ is the number of capacitors that sample ΔV_{BE} (see Fig. 6.5). During Φ_2, the input voltages are inversed and charges equal to $C_S\cdot K_i\cdot(\Delta V_{BE})$ and $C_S\cdot 1\cdot(-V_{BE})$ are sampled. The resulting voltage at the amplifier's output $2C_S/C_{I1}\cdot(V_{BE}-K_i\cdot V_{BE})$ is then evaluated by the comparator, which indicates whether V_{BE} is larger or smaller than $K_i\cdot\Delta V_{BE}$. This sampling and evaluation continues for five consecutive cycles and K_i of each cycles is updated with a binary-search logic. Eventually the range $[K_i(5th), K_i(5th)+1]$ will be found, which bounds X_T. In the sixth cycle, the SAR decides whether $K_i(5th)$ or $K_i(5th)+1$ is closer to X_T by comparing V_{BE} with $[K_i(5th)+0.5]\cdot\Delta V_{BE}$. To make this comparison, a charge equal to $C_S\cdot[1\cdot V_{BE}-K_i(5th)\cdot\Delta V_{BE}]$ is sampled in Φ_1, and a charge equal to $C_S\cdot[-1\cdot V_{BE}+(K_i(5th)+1)\cdot\Delta V_{BE}]$ is sampled in Φ_2, resulting in a voltage at the integrator's output equal to $2C_S/C_I\cdot[V_{BE}-(K_i(5th)+0.5)\cdot\Delta V_{BE}]$.

During the second step of the conversion, a $\Sigma\Delta$ modulator is realized around a 2nd-order loop filter. The first integrator samples and integrates a voltage equal to $2C_S/C_I\cdot[V_{BE}-(K_{SAR}+1)\cdot\Delta V_{BE}]$ when $bs=1$ and a voltage equal to $2C_S/C_I\cdot[V_{BE}-(K_{SAR}-1)\cdot\Delta V_{BE}]$ when $bs=0$. This is equivalent to having V_{BE} as the input signal, and $(K_{SAR}-1)\cdot\Delta V_{BE}$ and $(K_{SAR}+1)\cdot\Delta V_{BE}$ as the references for the $\Sigma\Delta$ modulator.

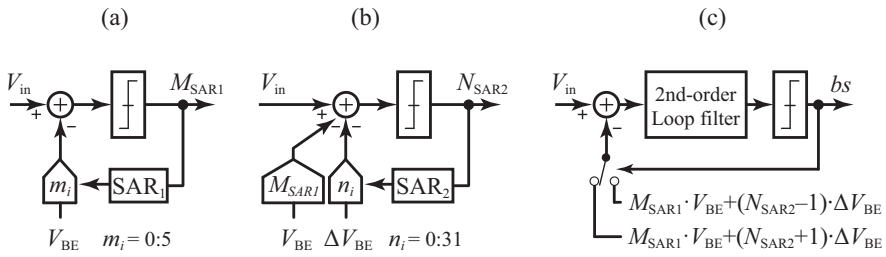


Figure 6.3: Three-step conversion in voltage sensing mode: (a) SAR1 finds M_{SAR1}, (b) SAR2 finds N_{SAR2} and (c) the $\Sigma\Delta$ modulator finds $\mu_{\Sigma\Delta}$.

6.3 Voltage-to-digital conversion

6.3.1 Principle of operation

One way to implement the voltage sensing mode is by reusing the zoom-ADC, replacing V_{BE} with V_{in}, and then digitizing $X_V = V_{in}/\Delta V_{BE}$. However, considering that the existing CDAC has 31 elements, and ΔV_{BE} is only $\sim$ 30 mV at -55°C, there would be an upper limit for the V_{in} to be digitized: $31 \cdot 30$ mV $\sim$ 0.93 V. To handle a full-scale V_{in} of 1.8 V, the number of CDAC elements needs to increase to $\sim$ 60, and the first integrator would then require a different OTA to charge a larger load capacitance. This means additional silicon area, more complexity in the OTA design, and larger leakage currents in the sampling switches. To avoid these drawbacks, and implement the voltage sensing mode with the same OTA and CDAC, a three-step conversion scheme is used as illustrated in Fig. 6.3. It consists of two SAR steps followed by a $\Sigma\Delta$ conversion. The first step, SAR1, digitizes $M_{SAR1} = V_{in}/V_{BE}$ and then a residual voltage $V_{in} - M_{SAR1} \cdot V_{BE}$ is passed on to a second step, SAR2, to be digitized by ΔV_{BE}. In the third step, a $\Sigma\Delta$ modulator is used, and as in the temperature sensing mode, reference signals are selected around the input voltage with a $2 \cdot \Delta V_{BE}$ range. This *zooming* principle helps to minimize the residual charge fed into the $\Sigma\Delta$ loop-filter and relaxes the OTA's settling requirements. A simplified operation of this three-step conversion is illustrated in Fig. 6.3 and Fig. 6.4.

During $SAR1$, V_{in} is digitized by V_{BE}. Since V_{BE} is larger than ΔV_{BE}, with a minimum of $\sim$ 0.4 V at 125 °C, V_{in}/V_{BE} can reach a maximum of 4.5. Therefore, SAR1 can find the range $[M_{SAR1}, M_{SAR1} + 1]$ which bounds V_{in}/V_{BE} in three cycles using a binary search (see Fig. 6.3 (a), and Fig. 6.4 (a)). During these three cycles, in Φ_1, a charge equal to $C_S \cdot V_{in}$ is sampled on a C_S that is separate from the CDAC

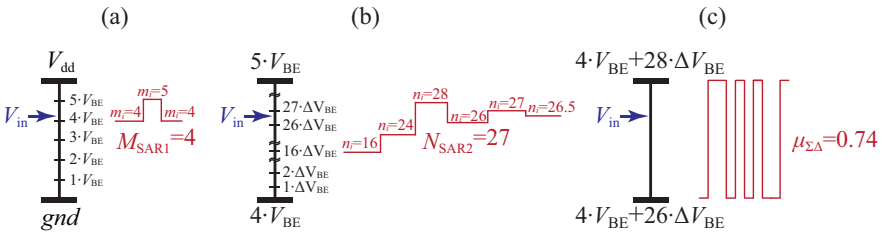


Figure 6.4: Simplified operation of the three-step conversion in voltage sensing mode: (a) SAR1, (b) SAR2 and (c) the ΣΔ modulator.

elements (see Fig. 6.5), and a charge equal to $m_i \cdot C_S \cdot (-V_{BE})$ is sampled on a set of m_i capacitors in the CDAC in differential branches, where $m_i = 1:5$. During Φ_2, charges equal to $C_S \cdot (-V_{in})$ and $m_i \cdot C_S \cdot V_{BE}$ are sampled, which then cause a voltage equal to $2C_S/C_{I1} \cdot (V_{in} - m_i \cdot V_{BE})$ at the first OTA output. Similar to the SAR conversion in the temperature sensing mode, the comparator and a binary search logic find M_{SAR1}.

In the second step of the conversion, SAR2, $M_{SAR1} \cdot V_{BE}$ is subtracted from V_{in}, and the residual voltage $V_{in} - M_{SAR1} \cdot V_{BE}$ is digitized by ΔV_{BE} (see Fig. 6.3 (b), and Fig. 6.4 (b)). This is done by sampling a charge equal to $C_S \cdot (V_{in} - M_{SAR1} \cdot V_{BE} - n_i \cdot \Delta V_{BE})$ in Φ_1, and a charge equal to $C_S \cdot (-V_{in} + M_{SAR1} \cdot V_{BE} + n_i \cdot \Delta V_{BE})$ in Φ_2, which then generates a voltage equal to $2C_S/C_{I1} \cdot (V_{in} - M_{SAR1} \cdot V_{BE} - n_i \cdot \Delta V_{BE})$ at the output of the first OTA, where $n_i = 1:31$ is the set of capacitors of the CDACs that sample ΔV_{BE}. Given that the residual voltage after SAR1 is lower than V_{BE}, the 31 elements of the CDAC are enough to digitize this voltage with ΔV_{BE} steps. SAR2 thus works exactly the same as the SAR conversion in the temperature sensing mode, and in six consecutive cycles finds the closest integer N_{SAR2} to $(V_{in} - M_{SAR1} \cdot V_{BE})/\Delta V_{BE}$.

In the third step of the conversion, a 2nd-order ΣΔ modulator balances V_{in} against $M_{SAR1} \cdot V_{BE} + (N_{SAR2} - 1) \cdot \Delta V_{BE}$ and $M_{SAR1} \cdot V_{BE} + (N_{SAR2} + 1) \cdot \Delta V_{BE}$. This is done by sampling and integrating a voltage equal to $2C_S/C_{I1} \cdot V_{in}$ and a voltage equal to $2C_S/C_{I1} \cdot [M_{SAR1} \cdot V_{BE} + (N_{SAR2} - 1) \cdot \Delta V_{BE}]$ if $bs = 0$, or a voltage equal to $2C_S/C_{I1} \cdot [M_{SAR1} \cdot V_{BE} + (N_{SAR2} + 1) \cdot \Delta V_{BE}]$ if $bs = 1$. This results in a bs whose average, $\mu_{\Sigma\Delta}$, satisfies the following:

$$\begin{aligned}(\mu_{\Sigma\Delta}) \cdot [V_{in} - (M_{SAR1} \cdot V_{BE}) - (N_{SAR2} + 1) \cdot \Delta V_{BE}] \\ +(1 - \mu_{\Sigma\Delta}) \cdot [V_{in} - (M_{SAR1} \cdot V_{BE}) - (N_{SAR2} - 1) \cdot \Delta V_{BE}] = 0\,,\end{aligned} \tag{6.1}$$

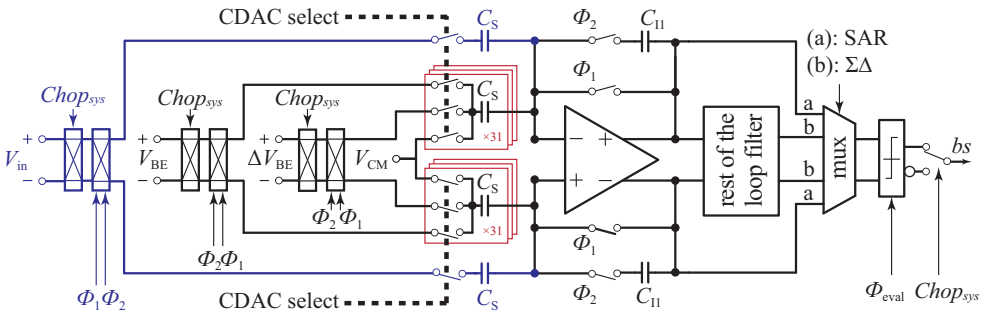


Figure 6.5: Schematic of the readout circuit in voltage sensing mode. (Black) reused from the zoom-ADC and (blue) additional circuitry.

which leads to:

$$\frac{V_{in}}{\Delta V_{BE}} = X_V = M_{SAR1} \cdot \frac{V_{BE}}{\Delta V_{BE}} + (N_{SAR2} - 1) + 2 \cdot \mu_{\Sigma\Delta}$$
$$= M_{SAR1} \cdot X_T + (N_{SAR2} - 1) + 2 \cdot \mu_{\Sigma\Delta}\,, \tag{6.2}$$

where X_T is the untrimmed output of the temperature-to-digitial conversion.

To find the absolute value of V_{in}, ΔV_{BE} can be replaced by its absolute value. This can be done using the trimmed output of the temperature-to-digital converter and a linear mapping function as:

$$V_{in} = X_V \cdot [C \cdot (D_{T,out} - 25°C) + D]\,, \tag{6.3}$$

where $D_{T,out}$ is the trimmed output of the temperature sensing mode, C ($\sim$ 140 μV/C) is a calibration parameter representing the temperature dependence of ΔV_{BE}, and D ($\sim$ 42 mV) is its absolute value at 25 °C. As shown in Chapter 4, since ΔV_{BE} is quite stable, C and D can be found with a one-time calibration. It is also worth noting that the curvature of V_{BE} is mitigated in the temperature sensing mode, and so only appears as a residual non-linearity in the ΔV_{BE} estimation. Therefore, no extra V_{BE} curvature correction is required.

6.3.2 Accuracy of the voltage sensing mode

The accuracy of the voltage sensing mode is limited by the spread of ΔV_{BE}, as described in (6.2). The temperature sensing accuracy $D_{T,out}$ also sets another limit on the voltage-sensing accuracy, since it is used to estimate ΔV_{BE} in (6.3). Furthermore,

readout errors in X_V add up to the voltage sensing error. V_{in} is sampled on a separate capacitor C_S, rather than the dynamically matched capacitors of the CDAC array. This causes a gain-error on the voltage sensing. However, due to their low temperature coefficients ($\sim$ 15 ppm/°C), the mismatch between capacitors is a weak function of temperature, and so this error is fixed and can be corrected by a single point calibration at room temperature.

In addition to these error sources, the sampling switches that connect V_{in} to C_S, in this design, are driven by normal logic gates supplied by the power supply rails, which in turn causes the sampled voltage to become a function of the on-resistance R_{on} of the switches. This creates gain errors in the voltage sensing modes since R_{on} depends on V_{in}, and so creates input-dependent sampling errors.

The accuracy of the voltage sensing mode could be improved by using one of the elements in the dynamically matched capacitor arrays to sample V_{in}. This, however, would cause more leakage as there would be 64 more switches connected to the CDAC. Gain errors caused by the input-dependent R_{on} of the sampling switches could also be suppressed by using a bootstrapped driving circuit as suggested by [6.5]. Using these methods would result in a voltage sensing accuracy that is only limited by the DC accuracy of ΔV_{BE}, and the temperature-sensing accuracy $D_{T,out}$.

6.4 Capacitance-to-digital conversion

As shown in Fig. 6.6, in the capacitance sensing mode, the zoom-ADC is again reused and a similar two-step conversion approach adopted. The difference is that the sampled charge is now proportional to C_{in} and C_S instead of two temperature-dependent voltages V_{BE} and ΔV_{BE}. An external capacitance C_{in} is first digitized using a SAR with respect to C_S, to find the closest integer possible L_{SAR} to $X_C = C_{in}/C_S$, where C_S is the unit capacitance of the CDAC array. A modulator then balances C_{in} against $(L_{SAR} - 1) \cdot C_S$ and $(L_{SAR} + 1) \cdot C_S$.

In the first step, SAR, the first integrator is configured as a charge amplifier (see Fig. 6.7) and together with the comparator and a SAR logic finds L_{SAR}. In a 6-cycle SAR, during Φ_1, a charge equal to $C_{in} \cdot V_{dd}$ is sampled on C_{in} in the differential branches, and a charge equal to $l_i \cdot C_S \cdot (-V_{dd})$ is sampled on sets of l_i capacitors of the CDAC, where $l_i = 1 : 31$. During Φ_2, the sampling voltage is inversed, and so charges equal to $C_{in} \cdot (-V_{dd})$ and $l_i \cdot C_S \cdot (V_{dd})$ are sampled. At the end of Φ_2, the difference between the sampled charges in Φ_1 and Φ_2 creates an output voltage equal to $2(C_{in} - l_i \cdot C_S)/C_{I1} \cdot V_{dd}$, whose polarity is evaluated by the comparator.

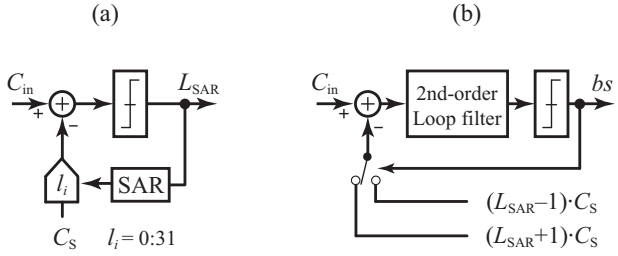


Figure 6.6: Two-step conversion in capacitance sensing mode: (a) SAR finds L_{SAR} as the closest integer to X_C and (b) $\Sigma\Delta$ finds $\mu_{\Sigma\Delta}$.

This procedure of sampling and comparison continues for six cycles. L_{SAR} is found in exactly same fashion as in the temperature sensing mode.

In the second step, a 2nd-order $\Sigma\Delta$ modulator is implemented using the 1st and the 2nd stage integrators and a passive adder that together form the loop-filter (see Fig. 6.7). A voltage equal to $2[C_{in} - (L_{SAR} - 1) \cdot C_S]/C_{I1} \cdot V_{dd}$ is integrated when $bs = 0$, or a voltage equal to $2[C_{in} - (L_{SAR} + 1) \cdot C_S]/C_{I1} \cdot V_{dd}$ when $bs = 1$. This scheme then creates a bs whose average $\mu_{\Sigma\Delta}$ can be found using:

$$\begin{aligned}(\mu_{\Sigma\Delta}) \cdot [C_{in} - (L_{SAR} + 1) \cdot C_S] \\ +(1 - \mu_{\Sigma\Delta}) \cdot [C_{in} - (L_{SAR} - 1) \cdot C_S] = 0\,,\end{aligned} \tag{6.4}$$

This equation can then be further simplified to find the value of C_{in} as:

$$\frac{C_{in}}{C_S} = X_C = (L_{SAR} - 1) + 2 \cdot \mu_{\Sigma\Delta}\,. \tag{6.5}$$

In the capacitance sensing mode, the sampled and integrated charges during the $\Sigma\Delta$ modulation are larger than those in the temperature and voltage sensing modes. This is because capacitors sample V_{dd} for better noise performance (see Sec. 6.5.3) compared to V_{BE} and ΔV_{BE} in the other two sensing modes. This creates a larger voltage at the output of the first integrator: $2V_{dd}\cdot(C_S/C_{I1})$ versus $2\Delta V_{BE}\cdot(C_S/C_{I1})$, which is about 40× higher at room temperature. In order to bring this output swing down to an acceptable range for the OTA, as shown in Fig. 6.7, an extra capacitance C_{extra} (= 2 pF) is added to C_{I1} (= 240 fF) in the capacitance sensing mode. In addition, to be able to test the performance of the capacitance sensing mode without an available external capacitor, an internal test capacitance C_{test} is realized which can be activated by en_{Ctest}.

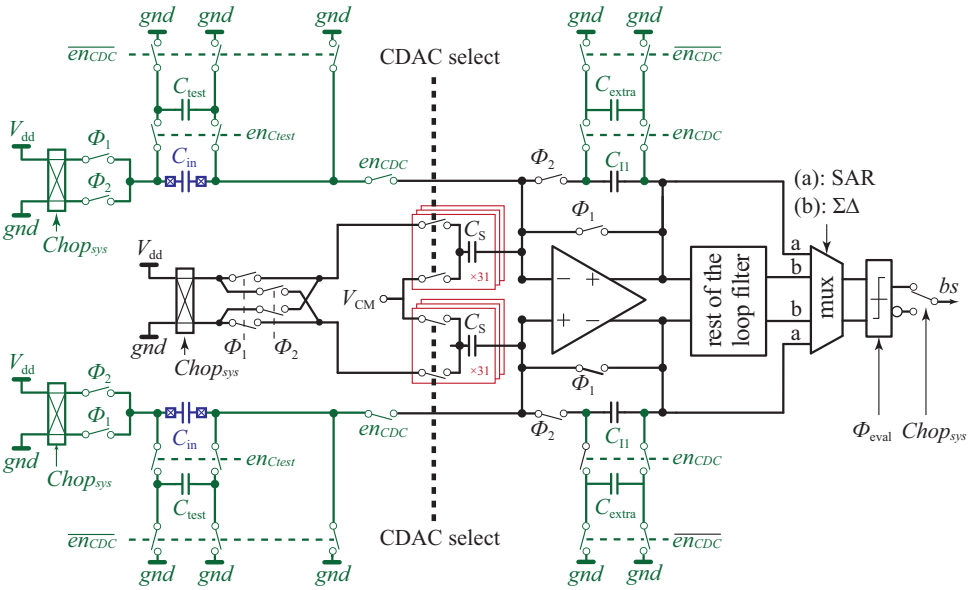


Figure 6.7: Schematic of the readout circuit in capacitance sensing mode. (Black) reused from the zoom-ADC, (green) additional circuitry and (blue) input capacitance.

6.5 Noise analysis

6.5.1 Noise in the temperature sensing mode

Noise in the temperature sensing mode originates from the noise in the front-end signals V_{BE} and ΔV_{BE} and from the readout circuitry. In [6.6], a detailed analysis of noise has been provided for TDCs with switched-capacitor readout circuitry. It has been shown that the contribution of noise from the switched-capacitor readout with $C_S = 120fF$ is substantially higher than the noise of the rest of the readout circuitry. This is mainly due to the noise-aliasing in switched-capacitor circuits. Using a continuous-time readout circuit can help reduce the TDC noise and obtain substantially higher resolution and also energy-efficiency FoM [6.7]. In this section, the noise in the temperature sensing is evaluated by taking into account only the major contributing sources, i.e., the switched-capacitor readout circuit.

As described in Section 6.2, errors in the SAR conversion will be absorbed into the value of $\mu_{\Sigma\Delta}$ found in the $\Sigma\Delta$ conversion. The noise of the TDC is thus due to the noise of the $\Sigma\Delta$ conversion. In the first stage of the $\Sigma\Delta$ loop-filter, two capacitors in differential branches sample V_{BE}, and $2(K_{SAR}\pm 1)$ capacitors sample ΔV_{BE} based on the bs. This is done in two sampling phases: Φ_1 and Φ_2. Using the method described

in [6.8], during Φ_1, the noise of the sampling switches with a spectral density of $4kTR_{on}$ and the equivalent input-referred noise of the OTA with a spectral density of $(16/3)kT/g_m$ are sampled on C_S. The total sampled noise charge on each sampling capacitor can be found using:

$$\overline{q^2_{n\phi 1,C_S}} = kTC_S \cdot (1 + \frac{1/3}{1+x}), \tag{6.6}$$

where k is the Boltzmann constant, T is the absolute temperature in Kelvin and C_S is the value of the sampling capacitor. $x = 2 \cdot g_m \cdot R_{on}$, where g_m is the transconductance of the OTA and R_{on} is the on-resistance of the sampling switch. This equation assumes that the OTA has an infinite output resistance R_L.

During Φ_2, the same noise charge as in Φ_1 will be sampled on each C_S, therefore, for each C_S the same total noise charge as (6.6) is obtained. It is worth noting that the first integrator in Fig. 6.5 uses correlated double sampling, and unlike the stray-insensitive integrator described in [6.8], it samples the same noise charge in both phases. The total noise charge on C_S will then be the sum of noise charges in each phase. In most cases, and in particular this design, for the lowest power consumption, the OTA is the main element which limits the settling time of the charge sampling. This means that $g_m \gg 1/R_{on}$ or $x \ll 1$ can be assumed. By noting this and using (6.6), the total sampled noise charge on C_S during one $\Sigma\Delta$ cycle can be expressed as:

$$\overline{q^2_{n,C_S}} = 2 \cdot \overline{q^2_{n\phi 1,C_S}} = 2 \cdot kTC_S \cdot (1 + \frac{1/3}{1+2x}) \approx 2 \cdot (\frac{4}{3}) \cdot kTC_S\,. \tag{6.7}$$

Since this sampled noise charges are uncorrelated for each sampling capacitor, the total noise charge that is sampled in one cycle of $\Sigma\Delta$ can be found using $\overline{q^2_n}(n) = \overline{q^2_{n,V_{BE}}}(n) + \overline{q^2_{n,\Delta V_{BE}}}(n)$ where:

$$\overline{q^2_{n,Vbe}} = 2 \cdot \overline{q^2_{n,C_S}} \approx 2 \cdot (\frac{8}{3}) \cdot kTC_S\,, \tag{6.8}$$

$$\overline{q^2_{n,\Delta Vbe}} = 2(K_{SAR} \pm 1) \cdot \overline{q^2_{n,C_S}} \approx 2(K_{SAR} \pm 1) \cdot (\frac{8}{3}) \cdot kTC_S\,, \tag{6.9}$$

and the additional factor 2 account for the differential branches.

In each cycle of $\Sigma\Delta$ a signal charge equal to $V_{BE} \cdot 2C_S$ and $\Delta V_{BE} \cdot 2(K_{SAR} \pm 1) \cdot C_S$ is also sampled. Using a mathematical model of the $\Sigma\Delta$ loop-filter as shown in Fig. 6.8, the noise charges in (6.8) and (6.9) can thus be modelled as equivalent voltage

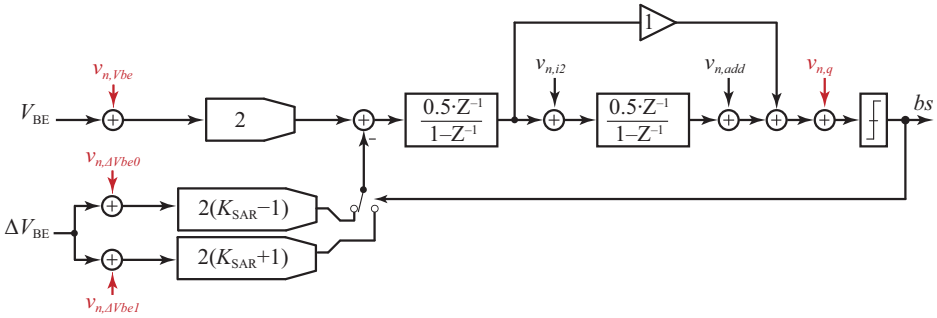


Figure 6.8: Model of the ΣΔ modualtor with equivalent noise sources of different stages. This model is used to evaluate the output noise in temperature sensing mode.

noise sources added to V_{BE} and ΔV_{BE} with rms values as:

$$v_{n,Vbe} = \frac{\sqrt{\overline{q_{n,Vbe}^2}}}{2C_S} \approx \sqrt{\frac{4 \cdot kT}{3 \cdot C_S}}, \tag{6.10}$$

$$v_{n,\Delta Vbe} = \frac{\sqrt{\overline{q_{n,\Delta Vbe}^2}}}{2(K_{SAR} \pm 1)C_S} \approx \sqrt{\frac{4 \cdot kT}{3(K_{SAR} \pm 1) \cdot C_S}}. \tag{6.11}$$

The rest of the CDAC capacitors are not considered to be noise contributors. This is valid since the unused capacitors can be shorted to exclude their noise contribution; an assumption that is made for the noise analysis in this work. The input-referred noise of the other stages of the loop-filter such as the 2nd-stage integrator and the passive adder can also be found in the same way as (6.10) and (6.11). They can then be added to their inputs in Fig. 6.8.

The next step in the noise evaluation process is to account for the noise transfer function of each noise source, first to the output of the ADC, denoted as X_T, and then to the temperature output, $D_{T,out}$. This can be done analytically, as described in [6.8], or by using a numerical simulator such as MATLAB. In either case, it can be shown that the input noise contribution at the 2nd-stage integrator ($V_{n,i2}$) and the passive adder ($V_{n,add}$) to the output would be significantly lower than at the 1st-stage integrator. This reduction is due to the shaping and subsequent suppression of these noise components by the loop filter.

For the 2nd-order modulator shown in Fig. 6.8 and operating at an oversampling ratio of OSR, $V_{n,i2}$ is 1st-order shaped (suppressed by $\sim \sqrt{OSR^3}$), while $V_{n,add}$ is 2nd-order shaped (suppressed by $\sim \sqrt{OSR^5}$). The OSR is equivalent to the ratio of the sampling frequency f_S to twice the bandwidth of the desired signal f_{BW} in

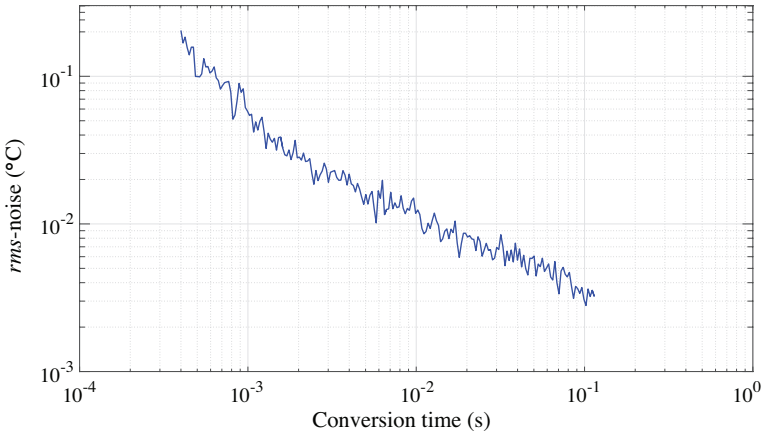


Figure 6.9: Simulated output noise in temperature sensing mode using the model in Fig. 6.8.

a $\Sigma\Delta$ modulator. In an incremental modulator, the signal is DC, allowing for an arbitrary signal bandwidth. Consequently, the OSR equates to the number of $\Sigma\Delta$ cycles required for one conversion, which can be calculated as $conversion\ time \times f_S$. With $f_S = 35$ kHz and a conversion time of 100 ms, $V_{n,i2}$ is suppressed by a factor of $\sim 10^5$ and $V_{n,add}$ by a factor of $\sim 10^8$, resulting in negligible contributions to the overall noise.

In the MATLAB model shown in Fig. 6.8, the noise corresponding to the 1st-stage integrator is modeled by a random number generator, with a Gaussian spectral distribution and standard deviation equal to (6.10) and (6.11). Fig. 6.9 illustrates the simulated output noise in temperature sensing mode at room temperature, resulting in an rms-noise of 3.5 mK in a conversion time of 100 ms. This output noise is calculated as the standard deviation of $D_{T,out}$ resulting from the modulator after repeating it 100 times. $D_{T,out}$ in each repetition is found from $D_{T,out} = A \cdot \alpha/(\alpha + X_T) - B$, where $X_T = K_{SAR} + 2 \cdot \mu_{\Sigma\Delta}$ is obtained from the SAR output and the $sinc^2$-decimated bs of the modulator. To show the noise performance for different conversion times, the same procedure is repeated but with different conversion time values, showing a good correlation with the measured result of the TDC in Fig. 4.17 (a).

6.5.2 Noise in the voltage sensing mode

Noise in the voltage sensing mode consists of three noise components: the noise of the $D_{T,out}$ in the 2nd term of (6.4), X_T in the 1st term of (6.4), as well as $\mu_{\Sigma\Delta}$ in the

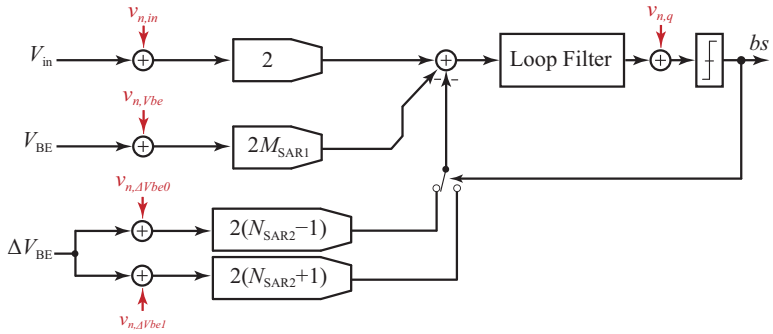


Figure 6.10: Model of the ΣΔ modualtor with the equivalent noise sources. This model is used to evaluate the output noise in voltage sensing mode.

first term of (6.4). To find the noise in $\mu_{\Sigma\Delta}$, in a similar way to temperature sensing mode, the sampled noise charge is found and then an equivalent *rms* voltage noise is added to the mathematical model of the ΣΔ modulator as illustrated in Fig. 6.10. Similar to (6.10) and (6.11), the equivalent *rms* noise voltages that adds to V_{in}, and ΔV_{BE} can be expressed as:

$$v_{n,in} \approx \sqrt{\frac{4 \cdot kT}{3 \cdot C_S}}, \tag{6.12}$$

$$v_{n,\Delta Vbe} \approx \sqrt{\frac{4 \cdot kT}{3 \cdot (N_{SAR2} \pm 1) \cdot C_S}}. \tag{6.13}$$

Fig. 6.11 illustrates the output noise of the voltage sensing mode, achieving an *rms* value of 11 μV in a conversion time of 100 ms. In this figure, the noise in $D_{T,out}$ and X_T is taken into account from the temperature sensing mode. This result correlates well with the measured results that will be provided in Section 6.6.

6.5.3 Noise in the capacitance sensing mode

In the capacitance sensing mode, the noise originates from the sampled noise on the input capacitance C_{in} and the capacitances of the CDAC. The first stage of the modulator samples a charge equivalent to $C_{in} \cdot (2V_{dd})$ in each ΣΔ cycle. Simultaneously, a total sampled noise charge on C_{in} equals to $\overline{q^2_{n,C_{in}}}(n) = 2 \cdot (8/3) \cdot kT \cdot C_{in}$. As shown in the model of the modulator in Fig. 6.12, this noise is modelled with an equivalent

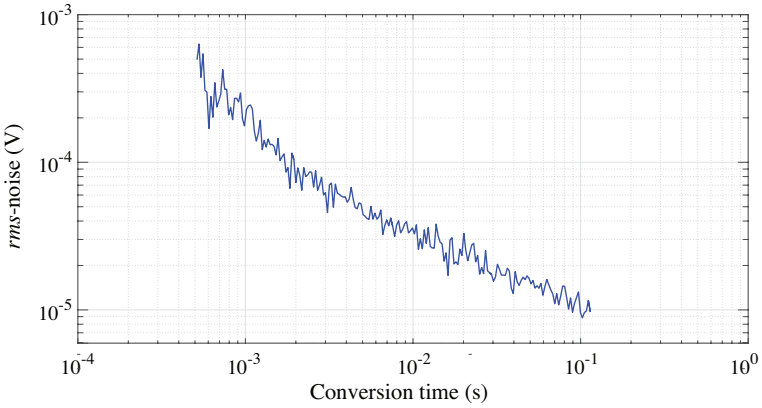


Figure 6.11: Simulated output noise in voltage sensing mode using the model in Fig. 6.10.

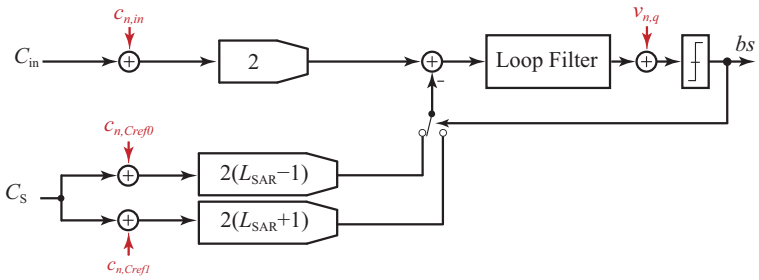


Figure 6.12: Model of the ΣΔ modualtor with the equivalent noise sources. This model is used to evaluate the output noise in capacitance sensing mode.

input noise capacitance with an *rms* value as:

$$c_{n,in} = \frac{\sqrt{\overline{q^2_{n,Cin}}}}{2V_{dd}} \approx \sqrt{\frac{4 \cdot kT \cdot C_{in}}{3 \cdot V^2_{dd}}}. \tag{6.14}$$

Similar to the input capacitance, for the capacitances of the CDAC array, a charge equivalent to $C_S \cdot (2V_{dd}) \cdot (L_{SAR} \pm 1)$ is sampled in each cycles, while a total noise charge of $\overline{q^2_{n,C_{ref}}}(n) = 2 \cdot (8/3) \cdot kT \cdot C_S \cdot (L_{SAR} \pm 1)$ is also sampled. This is then added as an equivalent input noise capacitance to the modulator with an *rms* value as:

$$c_{n,Cref} = \frac{\sqrt{\overline{q^2_{n,Cs}}}}{2V_{dd}(L_{SAR\pm1})} \approx \sqrt{\frac{4 \cdot kT \cdot C_S}{3 \cdot V^2_{dd} \cdot (L_{SAR} \pm 1)}}. \tag{6.15}$$

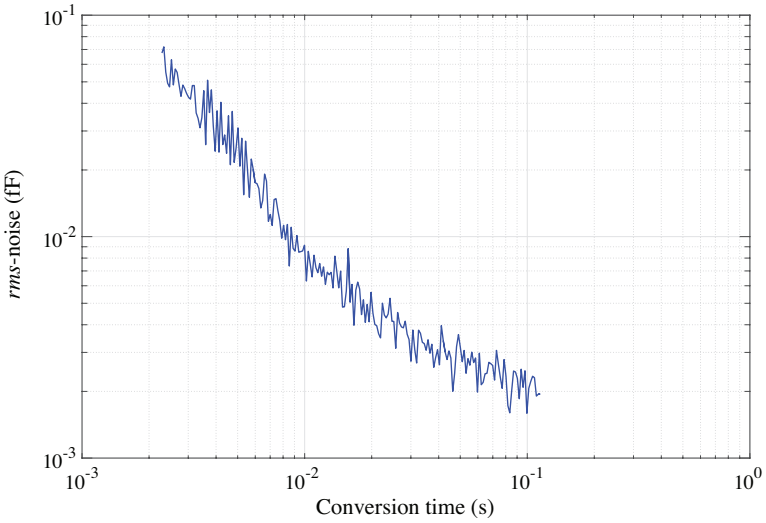


Figure 6.13: Simulated output noise in capacitance sensing mode using the model in Fig. 6.12.

Fig. 6.13 shows the simulated noise of the capacitance sensing mode, which obtains an rms noise of 2 aF in a conversion time of 100 ms.

6.6 Experimental results

The readout circuit was realized in 0.16-μm CMOS technology (see Fig. 6.14), occupying a total die area of 0.33 mm^2, and drawing 4.6 μA from a 1.8 V supply. The front-end and the ADC were the same as those used in the TDC described in Chapter 4. However, extra integration capacitors ($C_{I1,extra}$) were added to obtain an acceptable output swing in the capacitance sensing mode. Furthermore, on-chip test capacitors (C_{test}) were added to enable testing of the capacitance sensing mode.

The readout circuit was initially characterized in the TDC mode. Like the TDC in Chapter 4, it obtained a thermal noise limited resolution of 4 mK in a 100 ms conversion time. 24 samples of the chip were packaged in ceramic and then characterized from -55 °C to $+125$ °C. After a batch calibration to extract the parameters A, B, and α and without trimming, the TDC obtained ± 0.6 °C (3σ) error as shown in Fig. 6.16 (a). After temperature calibration and trimming at 30 °C, the errors reduced to ± 0.2 °C (3σ) as shown in Fig. 6.16 (b). This is about $2\times$ worse than the accuracy of TDC in Chapter 4, which is possibly due to the higher leakage caused by the extra switches that were added to handle the voltage and capacitance sensing

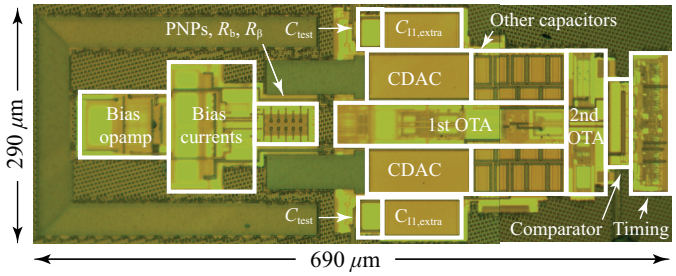


Figure 6.14: Chip micrograph of the readout circuit.

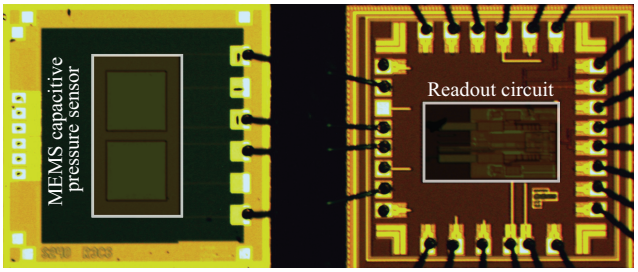


Figure 6.15: Connecting a dual-capacitance pressure sensor to the readout circuit.

modes. During the batch calibration, the parameters C and D were also found, which were then used to estimate ΔV_{BE} from $D_{T,out}$.

In the voltage sensing mode, the readout circuit obtained 12 μV_{rms} output noise in a conversion time of 100 ms, as illustrated in Fig. 6.17. The accuracy of the voltage sensing mode was limited by the accuracy of the temperature sensing mode (± 0.2 °C) used to estimate ΔV_{BE}, and the spread of ΔV_{BE} itself. The estimation accuracy of ΔV_{BE} is shown in Fig. 6.18 (a), which is about $\pm 0.1\%$. To obtain this estimation error, the measured ΔV_{BE} of each sample was compared to its estimation found by $C \cdot (D_{T,out} - 25\ °C) + D$. Fig. 6.18 (b) shows that the readout circuit achieves $\pm 0.5\%$ relative accuracy up to 1.8 V. The output results show a temperature-independent spread from sample-to-sample, which mainly originated from the spread in ΔV_{BE}, and also from the mismatches between C_S, which samples V_{in}, and the average unit capacitor of the CDAC. For the low input voltages (0.3 V), the results show a shift which was due to the signal-dependent errors caused by the sampling switches which were driven at the supply rails.

In the capacitance sensing mode, as shown in Fig. 6.19, the readout circuit obtained a thermal-noise limited resolution of 2 aF in a 100 ms conversion time corresponding to a resolution FoM of 0.76 pJ/step. To show the results of the readout cir-

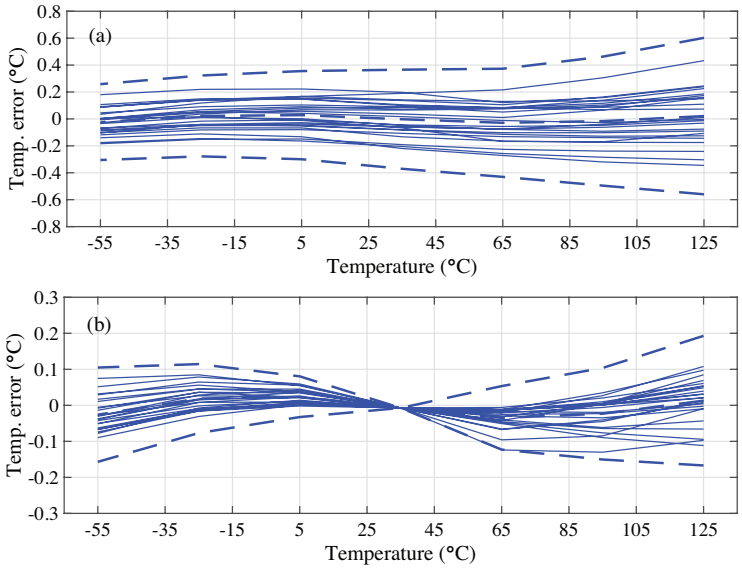


Figure 6.16: Inaccuracy of temperature sensing. Dashed lines represent the average and $\pm 3\sigma$ boundaries: (a) untrimmed and (b) after temperature calibration at 30°C.

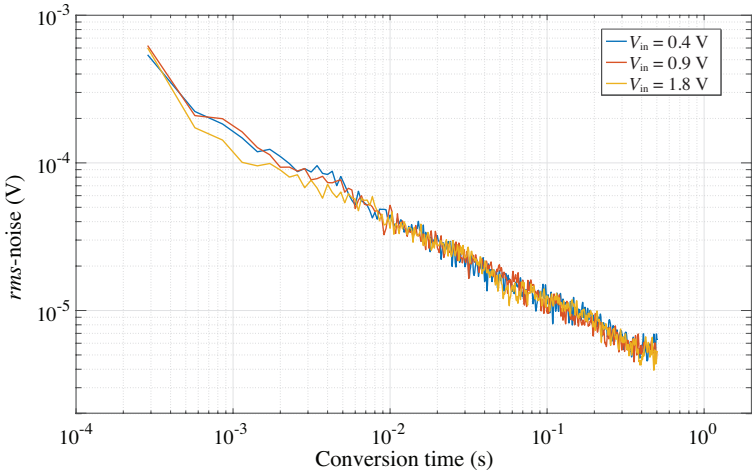


Figure 6.17: Measured output noise in voltage sensing mode.

cuit with a variable external capacitance, a few samples of the circuit were connected to a capacitive MEMS pressure sensor, as shown in Fig. 6.15, and then characterized under varying levels of ambient pressure. The output capacitance of the MEMS sensor changed in response to a change in ambient pressure essentially by changing the distance between a silicon diaphragm as one plate of the capacitor and a fixed

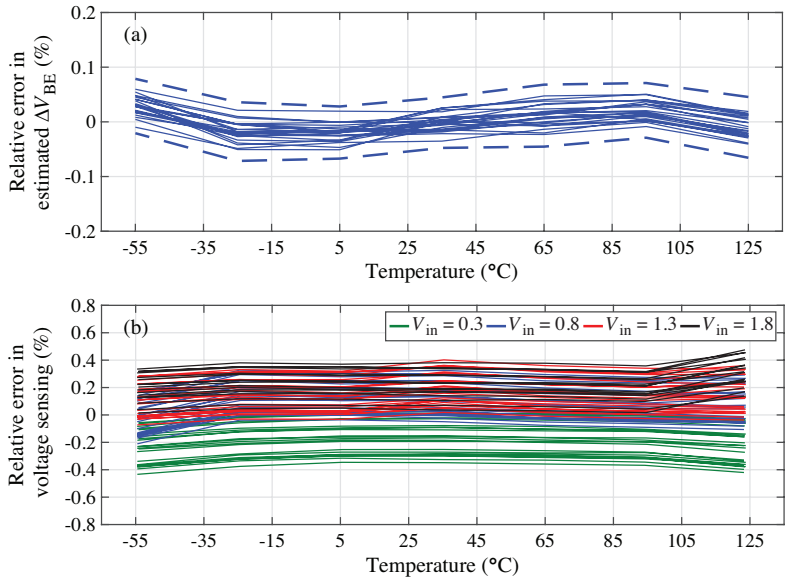


Figure 6.18: Measured inaccuracy of voltage sensing mode: (a) estimation error of ΔV_{BE} using temperature sensing output and (b) absolute inaccuracy in voltage sensing mode.

substrate as its second plate. The MEMS sensor used for the measurement contained two identical capacitive sensors whose output capacitance changed from $\sim$ 500 fF at 0 bar to $\sim$ 1.5 pF at 1.4 bar [6.9]. These two capacitors were then connected to the readout circuit in the differential branches.

The output of the sensor was observed in the presence of a slow but uncontrolled drift of pressure obtained by allowing the pressure of a vacuum environment to reach the room pressure using a relief valve. The output of the capacitance sensing is shown in Fig. 6.20 (a). To cover a different range of capacitance variation, the test capacitance (= 2 pF) was added to that of the MEMS sensor, while the same measurement was performed. The result is reported in Fig. 6.20 (b), showing that readout was capable of capturing continuous capacitance variation up to 3.1 pF.

Table 6.1 summarizes the performance characteristics of the readout circuit in the three different modes and then compares them to other state-of-the-art sensors with the combined temperature and voltage sensing functionalities, as well as the state-of-the-art stand-alone capacitance-to-digital converters. The readout circuit in this work [6.10] shows an enhanced performance in terms of energy efficiency compared to multi-functional readout circuits [6.11]–[6.13], while favourably compares with state-of-the-art stand-alone temperature sensors (Chapter 4) and capacitance sensors [6.14]–[6.16].

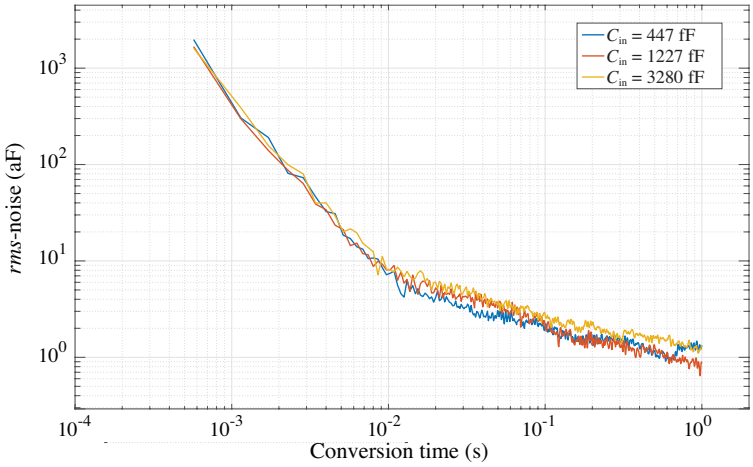


Figure 6.19: Measured output noise in capacitance sensing mode.

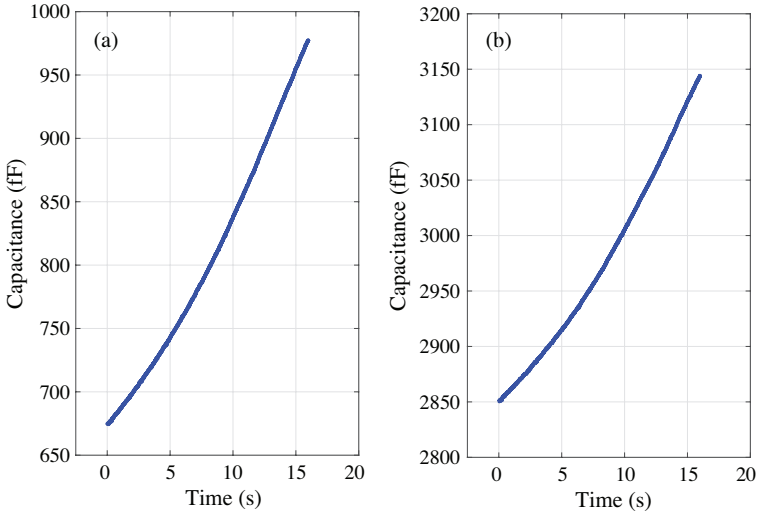


Figure 6.20: Measured capacitance of the MEMS pressure sensor. The pressure slowly drifts from vacuum to the room-pressure while the readout continuously captures its capacitance. (a) $C_{in} = C_{MEMS}$ and (b) $C_{in} = C_{MEMS} + (\sim 2\text{pF})$.

6.7 Conclusion

This chapter demonstrated the realization of a compact multi-function sensor, which can sense temperature, voltage and capacitance. This is obtained by reusing the front-end and the zoom-ADC of a high accuracy TDC. Such a compact realization of multiple functionalities can help reduce area-related costs in IoT sensing applications.

Item		**[6.10]**	[6.11]	[6.12]	[6.13]	[6.14]	[6.15]	[6.16]
Conversion method		**SAR-ΣΔ**	ΣΔ	ΣΔ	ΣΔ	SAR-ΣΔ	ΣΔ	I. D.[a]
Technology (μm)		**0.16**	0.18	0.7	N/A	0.18	0.16	0.04
Area (mm^2)		**0.33**	0.18	4.8	N/A	0.46	0.28	0.002
Supply voltage (V)		**1.5–2**	1	2.5–5.5	2.7–5.2	1.4	1.2	1
Supply current (μA)		**4.6**	0.72	85	500	24	8.6	1.84
Conversion time (ms)		**100**	40	100	200	0.23	0.8	0.019
Temp. sensing	Range (°C)	**−55 to 125**	0 to 100	−40 to 105	−40 to 125	-	-	-
	Res. (mK)	**4**	40	9	N/A	-	-	-
	FoM[b](pJK^2)	**11**	46	1875	N/A	-	-	-
	±3σ-IA (°C)	**±0.2**	±0.18	N/A	±0.5	-	-	-
Voltage sensing	Range (V_{pp})	1.8	-	5	5	-	-	-
	Rel. IA[c](%)	**±0.5**	-	±0.012[d]	±0.1	-	-	-
Cap. sensing	Range (pF)	**0-3.8**	0-10.8	-	0-8	0-24	-	0-1×10^3
	Res. (aF)	**2.5**	170	-	20	156	70	12×10^3
	ENOB[e]	**18.7**	14.1	-	19.1	15.4	12.5	8
	FoM[f]($pJ/step$)	**0.76**	1.3	-	6700	0.18	1.4	0.14

[a] Iterative discharge
[b] (Power×Conversion time)×Res.^{2}
[c] 100×Inaccuracy/Range
[d] Temperature drift of one sample
[e] $(1/6.02) \cdot (20 \cdot log_{10}(range/2 \cdot \sqrt{2}) - 1.76)$
[f] (Power×Conversion time)/2^{ENOB}

Table 6.1: Performance summary of the readout circuit in this work [6.10] and comparison with other related work.

In temperature sensing mode, it approaches the performance of the original TDC, with only slight degradation of its accuracy. In voltage sensing mode, the sensor is able to directly digitize full-scale (1.8 V) external voltages over the military temperature range. Instead of requiring external reference voltages for this operation, it uses its own V_{BE} and ΔV_{BE}, and obtains a voltage sensing accuracy of $\sim \pm 0.5\%$. The readout circuit can also digitize external capacitances of up to 3.8 pF with 2 aF resolution (i.e., 18.7 ENOB).

Bibliography

[6.1] C. Lu, S. Ravikumar, A. D. Sali, M. Eberlein, and H. Lee, "An 8b subthreshold hybrid thermal sensor with ±1.07°C inaccuracy and single-element remote-sensing technique in 22nm FinFET," in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, Feb. 2018, pp. 318–320.

[6.2] O. Bass and J. Shor, "Ultra-miniature 0.003 mm^2 PNP-based thermal sensor for CPU thermal monitoring," in *ESSCIRC 2018 - IEEE 44th European Solid State Circuits Conference (ESSCIRC)*, Sep. 2018, pp. 334–337.

[6.3] M. Eberlein and H. Pretl, "A no-trim, scaling-friendly thermal sensor in 16nm FinFET using bulk diodes as sensing elements," *IEEE Solid-State Circuits Letters*, vol. 2, no. 9, pp. 63–66, Sep. 2019.

[6.4] H. Reinisch, M. Wiessflecker, S. Gruber, H. Unterassinger, G. Hofer, M. Klamminger, W. Pribyl, and G. Holweg, "A multifrequency passive sensing tag with on-chip temperature sensor and off-chip sensor interface using EPC HF and UHF RFID technology," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 12, pp. 3075–3088, Dec. 2011.

[6.5] A. M. Abo and P. R. Gray, "A 1.5-V, 10-bit, 14.3-MS/s CMOS pipeline analog-to-digital converter," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 5, pp. 599–606, May 1999.

[6.6] K. Souri and K. A. A. Makinwa, *Energy-Efficient Smart Tempearture Sensors in CMOS Technology*, 1st ed., ser. Analog Circuits and Signal Processing. Springer International Publishing AG, Jan. 2018.

[6.7] S. H. Shalmany, K. Souri, U. Sönmez, K. Souri, M. D'Urbino, S. Hussaini, D. Tauro, and S. Tabatabaei, "A 620 μW BJT-based temperature-to-digital converter with 0.65 mK resolution and FoM of 190 $fJ \cdot K^2$," in *2020 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2020, pp. 70–72.

[6.8] R. Schreier, J. Silva, J. Steensgaard, and G. C. Temes, "Design-oriented estimation of thermal noise in switched-capacitor circuits," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 11, pp. 2358–2368, Nov. 2005.

[6.9] L. Rajendran, *Design of an energy-efficient interface circuit for MEMS-based pressure sensor*, ser. M.Sc. dissertation. Delft, the Netherlands: Delft University of Technology, 2013.

[6.10] B. Yousefzadeh, W. Wu, B. Buter, K. Makinwa, and M. Pertijs, "A compact sensor readout circuit with combined temperature, capacitance and voltage sensing functionality," in *2017 Symposium on VLSI Circuits*, June 2017, pp. C78–C79.

[6.11] T. Zhang, M. Law, P. Mak, M. Vai, and R. P. Martins, "Nano-watt class energy-efficient capacitive sensor interface with on-chip temperature drift compensation," *IEEE Sensors Journal*, vol. 18, no. 7, pp. 2870–2882, April 2018.

[6.12] N. Saputra, M. A. P. Pertijs, K. A. A. Makinwa, and J. H. Huijsing, "12-bit accurate voltage-sensing ADC with curvature-corrected dynamic reference," *Electronics Letters*, vol. 46, no. 6, pp. 397–398, March 2010.

[6.13] *AD7745 datasheet*, Analog Devices Inc., April 2005, rev. 0.

[6.14] Sechang Oh, Wanyeong Jung, Kaiyuan Yang, D. Blaauw, and D. Sylvester, "15.4b incremental sigma-delta capacitance-to-digital converter with zoom-in 9b asynchronous SAR," in *2014 Symposium on VLSI Circuits Digest of Technical Papers*, June 2014, pp. 1–2.

[6.15] Z. Tan, R. Daamen, A. Humbert, Y. V. Ponomarev, Y. Chae, and M. A. P. Pertijs, "A 1.2-V 8.3-nJ CMOS humidity sensor for RFID applications," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 10, pp. 2469–2477, Oct. 2013.

[6.16] W. Jung, S. Jeong, S. Oh, D. Sylvester, and D. Blaauw, "A 0.7pF-to-10nF fully digital capacitance-to-digital converter using iterative delay-chain discharge," in *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, Feb. 2015, pp. 1–3.

[6.17] J. Angevare, L. Pedalà, U. Sönmez, F. Sebastiano, and K. A. A. Makinwa, "A 2800-μm^2 thermal-diffusivity temperature sensor with VCO-based readout in 160-nm CMOS," in *2015 IEEE Asian Solid-State Circuits Conference (A-SSCC)*, Nov. 2015, pp. 1–4.

[6.18] U. Sönmez, F. Sebastiano, and K. A. A. Makinwa, "1650 μm^2 thermal-diffusivity sensors with inaccuracies down to $\pm$0.75°C in 40nm CMOS," in *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, Jan. 2016, pp. 206–207.

Chapter 7

Conclusions

This thesis explored methods for improving the accuracy and reducing the production costs of BJT-based temperature-to-digital converters (TDCs) in standard CMOS technology. This concluding chapter consists of two sections. Section 7.1 lists the main findings of this thesis, while Section 7.2 provides few suggestions for future work.

7.1 Main findings

- Within a batch, BJT-based TDCs in 0.16-μm CMOS technology can achieve an inaccuracy of $\pm$60mK (3σ) from -55°C to $+125$°C by designing sufficiently accurate readout electronics (Chapter 4).
- With sufficiently accurate readout electronics, the batch-to-batch spread of BJT-based TDCs in 0.16-μm CMOS technology can be reliably trimmed to within ±0.1°C (3σ) in the range from -55°C to $+125$°C on the basis of a one-point calibration (Chapter 4).
- Voltage calibration continues to be a reliable and low-cost alternative to temperature calibration in 0.16-μm CMOS technology, achieving an inaccuracy of ±0.3°C (3σ). This is due to the reproducibility of ΔV_{BE} across multiple batches and also in different packages (Chapters 4 and 5).
- When the mechanical stress of packaging is reproducible, its effect on a TDC can be separated from that of processing spread. As a result, a single-point calibration at wafer-level can correct the spread-related errors, while package-specific calibration can correct the stress-related errors (Chapter 5).

- If leakage-related errors are sufficiently suppressed, BJT-based sensors realized in a 0.16-μm technology can operate up to +180 °C with a 3σ error of ±0.2 °C (Chapter 5).
- Small on-chip heaters can locally elevate die temperature from ∼25°C to ∼100 °C in a fraction of a second. This elevated temperature, although not well-controlled, can be useful for calibration and trimming purposes (Chapter 5).
- Heater-assisted voltage calibration enables a low-cost and fast (< 0.5 s) two-point calibration without requiring any infrastructure except for an accurate voltage reference (Chapter 5).
- Heater-assisted bandgap trimming can correct BJT spread without the need for accurate external references. Although it is less accurate (3σ of ±1.4°C) than voltage (3σ of ±0.3°C) or temperature (3σ of ±0.1°C) calibration, it may be valuable for use in test environments where absolutely-accurate references are not available (Chapter 5).
- With a small overhead, the circuitry of a BJT-based TDC can be reused to sense external voltages or capacitances (Chapter 6).

7.2 Future work

Throughout this work, some ideas have been developed which can be useful in developing low-cost precision temperature sensors. Unfortunately, due to limited time, they only remained ideas and could not be realized or verified on silicon.

7.2.1 β-compensation for BJT-based TDCs

Voltage calibration [7.1] and the two heater-assisted calibration methods introduced in Chapter 5 rely on the intrinsic accuracy of ΔV_{BE}. While this accuracy has been validated in 0.16-μm CMOS technology, it may degrade in more advanced nodes due to the limited β values of available BJTs. The β-compensation method discussed in Chapter 4 [7.2] only corrects the impact of β on V_{BE} without affecting ΔV_{BE} (as explained in 4.4.2). The influence of β-related errors on ΔV_{BE}, as shown in the second term of equation (4.4), depends on both the absolute value of β and the flatness of the β curve with respect to current density. In cases where the β curve is not flat or β is low (e.g., < 1) [7.3],[7.4], an alternative β-compensation method can be used to mitigates the impact of β on ΔV_{BE}.

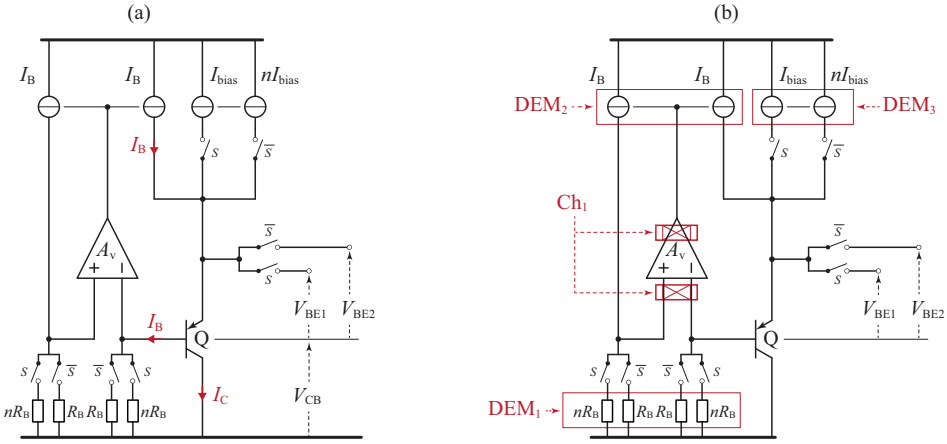


Figure 7.1: β-compensation. (a) operating principle, and (b) additional circuitry to remove non-idealities.

Figure 7.1(a) illustrates the operation of the proposed β-compensation scheme. It senses the base current (I_B) of the BJT (Q) and adds the same current to its emitter. Consequently, the collector current (I_C) is solely determined by the bias currents I_{bias} or $n \cdot I_{bias}$. To achieve this, a feedback loop comprising sense resistors R_B, an opamp A_V, and identical current mirrors is utilized. Depending on the state of S, either V_{BE1} or V_{BE2} is generated, allowing for subsequent sampling to obtain ΔV_{BE} ($= V_{BE2} - V_{BE1}$).

The accuracy of this β-compensation scheme is limited by the non-zero voltage across the collector-base junction (V_{CB}) of Q, which, due to the Early effect, results in non-PTAT errors in the saturation current (I_S). To reduce its effect on ΔV_{BE}, it is essential to keep V_{CB} the same at the two current densities. This is achieved by changing the value of the sense resistor from R_B to $n \cdot R_B$, when I_C changes from $n \cdot I_{bias}$ and I_{bias}, respectively . The residual error in ΔV_{BE} can then be derived as follows:

$$\Delta V_{BE} = \frac{kT}{q} \cdot ln(n) + \frac{kT}{q} \cdot \frac{\Delta\beta}{\beta^2} \cdot \frac{nI_{bias}R_B}{V_A}, \tag{7.1}$$

where V_A is the Early voltage, and $\Delta\beta$ is the difference in β at the two current densities. Assuming $n = 5$, $I_{bias} = 100nA$, $R_B = 50k\Omega$, and $V_A = 50V$, the residual error in ΔV_{BE} is reduced by approximately three orders of magnitude compared to the β-compensation method used in chapter 4.

Furthermore, the accuracy of the proposed β-compensation may be influenced by additional sources of error. Opamp offset, mismatches between sense resistors,

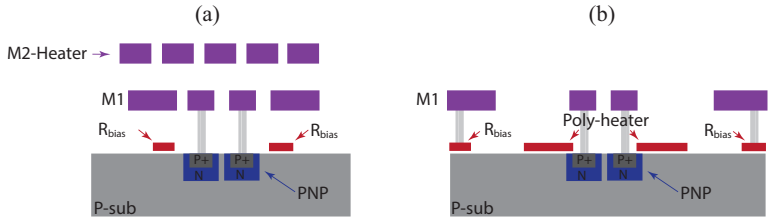


Figure 7.2: Illustration of heater position relative to PNPs and R_{Bias}: (a) Heater implementation from Chapter 5, and (b) More efficient heater using polysilicon.

mismatches between the components of the current mirrors for I_B, and inaccuracies in the ratio n between I_{bias} and $n \cdot I_{bias}$ can all contribute to overall errors. As illustrated in Fig. 7(b), these sources of error can be effectively mitigated through the application of techniques such as chopping (Ch_1) and dynamic element matching (DEM_1, DEM_2, and DEM_3).

7.2.2 Efficient on-chip heaters

The on-chip heater discussed in Chapter 5 can be optimized for increased efficiency in terms of power consumption. Utilizing polysilicon for the heater element allows it to be positioned much closer to the BJTs (see Figure 7.2), reducing its power requirement for achieving the same temperature rise. Lowering heater power consumption can also reduce the temperature rise in the surrounding sensor circuitry, thereby preventing calibration errors during local heating.

To implement this, the layout of the BJTs and the bias resistors must be adjusted to ensure uniform temperature elevation during local heating. While COMSOL simulations will provide a definitive solution, one preliminary idea involves relocating the polysilicon bias resistor further away from the BJTs compared to the layout shown in Chapter 5. Making connections to the BJTs, previously accomplished with Metal-1, presents another challenge. Depending on the heater's geometry and shape, using Metal-1 may still be a viable option.

Another issue that may be considered a challenge is the significantly higher sheet resistance of polysilicon, for example, being 1000× greater than that of metal. This may lead to a larger heater if both shape and resistance value must be maintained. . However, this challenge can be mitigated through several strategies: (a) using larger polysilicon strips, (b) implementing a parallel connection instead of series connection in a serpentine shape, and (c) driving the heater with a much lower current. It's important to note that the absolute value of the heater's resistance does not exclusively determine its power consumption, so maintaining the same resistance value may not be necessary.

Bibliography

[7.1] M. A. P. Pertijs, A. L. Aita, K. A. A. Makinwa, and J. H. Huijsing, "Low-cost calibration techniques for smart temperature sensors," *IEEE Sensors Journal*, vol. 10, no. 6, pp. 1098–1105, June 2010.

[7.2] M. Pertijs and J. Huijsing, "Bias circuits," Patent, Nov. 4, 2008, uS Patent 7,446,598 B2.

[7.3] M. Eberlein and H. Pretl, "Recent developments in bandgap references for nanometer CMOS technologies," in *2020 Austrochip Workshop on Microelectronics (Austrochip)*, 2020, pp. 42–46.

[7.4] Z. Tang, Y. Fang, Z. Huang, X. Yu, Z. Shi, and N. N. Tan, "An untrimmed BJT-based temperature sensor with dynamic current-gain compensation in 55-nm CMOS process," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 10, pp. 1613–1617, Oct 2019.

Appendix A

BJT theory

A.1 Ideal $I_C - V_{BE}$ characteristic

The collector current, I_C, of a PNP device is mainly determined by the minority carriers injected into the base from the emitter in the EB junction (I_{pE}), which diffuse across the base and are swept away by the collector in the CB junction (see Fig. A.1). To find I_C, requires finding the concentration of injected holes in the base p'. Since I_C is made from the diffusion of the holes, p' satisfies the *diffusion equation* as [A.1]:

$$\frac{d^2p'(x)}{dx^2} = \frac{p'(x)}{L_B^2}\,, \tag{A.1}$$

where x is the distance across the base width, and L_B is the diffusion length. $x = 0$ corresponds to the edge of the base in the EB junction, $x = W_B$ corresponds to the edge of the base in the CB junction, and W_B is the base width. L_B can be obtained from the recombination lifetime τ_B and diffusion constant D_B of the holes in the base as $L_B = (\tau_B D_B)^{0.5}$. The differential equation in (A.1) can be solved using the *Shockley boundary conditions*, which essentially define p' at the two edges of the base. These boundary conditions for p' at the left and the right edges of the base depend on V_{BE} and V_{BC} as [A.1]:

$$p'(0) = p_{B0} \cdot \left(exp(\frac{qV_{BE}}{kT}) - 1 \right)\,, \tag{A.2}$$

$$p'(W_B) = p_{B0} \cdot \left(exp(\frac{qV_{BC}}{kT}) - 1 \right) \approx 0\,, \tag{A.3}$$

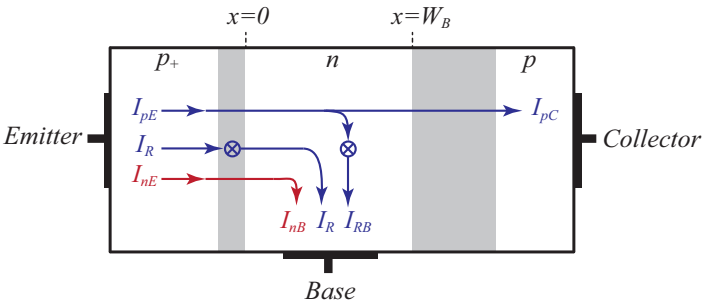


Figure A.1: Main current components in a PNP in the forward-active region.

where p_{B0} is the concentration of holes in the base at thermal equilibrium that follows $p_{B0} = n_i^2/N_B$, n_i is the intrinsic concentration of the base material (i.e., silicon), N_B is the base doping, k is the Boltzmann constant, q is the electron charge and T is the absolute temperature in Kelvin. A $p'(W_B)$ of zero can be assumed since all injected holes that reach their way to the edge of CB junction are immediately swept away by the collector. The solution of (A.1) can then be obtained by applying (A.2) and (A.3) as [A.1]:

$$p'(x) = \frac{n_i^2}{N_B} \cdot \left(exp(\frac{qV_{BE}}{kT}) - 1 \right) \cdot \left(\frac{sinh(\frac{W_B - x}{L_B})}{sinh(\frac{W_B}{L_B})} \right) . \tag{A.4}$$

L_B is typically a few tens of micro meters and is much larger than the usual W_B of actual devices, e.g., a maximum of a few hundreds of nanometers. Taking the assumption $L_B \gg W_B$ into account, and using the approximation $sinh(x) \approx x$ for $x \ll 1$, (A.4) can be expressed as [A.1]:

$$p'(x) \approx \frac{n_i^2}{N_B} \cdot \left(exp(\frac{qV_{BE}}{kT}) - 1 \right) \cdot (1 - \frac{x}{W_B}) . \tag{A.5}$$

Along the x axis in the base, the current density of the injected holes, J_B, can then be obtained using $J_B = -qD_B dp'/dx$, where the "$-$" sign denotes the holes being the current component. The collector current is thus found using $I_C = A_E J_B$, where the emitter area A_E is identical to the base area A_B. Taking these into account, I_C

can be expressed as [A.1]:

$$I_C = I_{pE} = -A_E q D_B \frac{dp'}{dx}$$

$$= A_E q \frac{D_B}{W_B} \frac{n_i^2}{N_B} \cdot \left(exp(\frac{qV_{BE}}{kT}) - 1 \right) . \tag{A.6}$$

Equation (A.6) expresses the ideal $I_C - V_{BE}$ characteristics in a PNP, which is usually expressed more compactly as:

$$I_C = I_S \cdot \left(exp(\frac{qV_{BE}}{kT}) - 1 \right) , \tag{A.7}$$

where I_S is known as the saturation current, which is expressed as:

$$I_S = A_E q \frac{D_B}{W_B} \frac{n_i^2}{N_B} . \tag{A.8}$$

A.2 Non-ideal current components

$\alpha_F = I_C/I_E$ in a PNP is equal to 1 when there is no base current involved. The Base current, however, is not zero and consists of three main components which are originating from different mechanisms. With reference to Fig. A.1, these current components can be described as the following:

(1) Minority carriers injected from the base to the emitter. In the forward-biased EB junction, minority carriers in the base can also be injected into the emitter (see Fig. A.1) in exactly the same manner as I_{pE}. This current can thus be expressed as [A.1]:

$$I_{nB} = A_E q \frac{D_E}{W_E} \frac{n_i^2}{N_E} \cdot \left(exp(\frac{qV_{BE}}{kT}) - 1 \right) . \tag{A.9}$$

where D_E, N_E, and W_E are, respectively, the diffusion constant of the electrons in the emitter, emitter doping and emitter width. This injected current does not contribute to the collector current but will equally add to the emitter current noted as I_{nE} $(= I_{nB})$ in Fig. A.1. As a result of this current, transistor behavior deviates from the ideal scenario causing $I_E \neq I_C$. To maximize α_F, and approach ideal transistor behavior, injection from the emitter to base I_{pE} needs to be substantially higher than injection from the base to emitter I_{nE}. In other words, γ, defined as the *emitter*

injection efficiency, needs to be maximized, which is expressed as:

$$\gamma = \frac{I_{pE}}{I_{pE} + I_{nB}} \,. \tag{A.10}$$

Replacing I_{pE} and I_{nB} in (A.10) by their values from (A.6) and (A.9), results in:

$$\gamma \approx \frac{1}{1 + \frac{N_B}{N_E} \cdot \frac{W_B}{W_E} \cdot \frac{D_E}{D_B}} \,. \tag{A.11}$$

To maximize γ, the emitter doping N_E needs to be substantially higher than N_B. Furthermore, W_B needs to be minimized compared to W_E. The former is almost always the case for a PNP device, while the latter depends on the geometry and type of the device. In vertical PNPs, W_B and W_E might not be very different since both are defined by the depth of a well or implanted impurity, which are not very well controlled, or at least not the to the same level as the accuracy of lithography. In these devices γ is then controlled by the ratio of the doping levels in the base and emitter regions.

(2) Recombination of injected minority carriers. Not all the injected holes into the base can make their way to the collector. This is due to the recombination of holes with the majority carriers that exist in the base. The amount of recombination depends on both the diffusion length L_B, and the base width W_B. As a result of this recombination effect, the collector current I_{pC} (see Fig. A.1) is lower than the injected emitter current I_{pE}. The ratio of I_{pC} to I_{pE} is defined as the *base transport factor*, α_T, which can be expressed as [A.2]:

$$\alpha_T = \frac{I_{pC}}{I_{pE}} \approx 1 - \frac{1}{2} \cdot (W_B/L_B)^2 \,. \tag{A.12}$$

α_T essentially determines the number of injected holes that could be transported through the base and reach the collector, hence the term *transport factor*. Reducing W_B can help to increase the α_T and approach ideal transistor behavior. The same recombination effect also happens in the emitter. In fact, it is fair to assume that nearly all the injected current in the emitter ($I_{nE} = I_{nB}$) is recombined. This is due to two reasons: firstly, the amount of injected minority carrier from the base is low due to its low doping, and secondly, due to the relatively higher doping of the emitter, the chance of being recombined with the majority carriers is high. Whether the carriers recombine in the emitter region or later on at the metal contacts of the emitter, this does not affect the collector current.

(3) Recombination-generation in the BE junction. Apart from the injection of minority carriers and recombination in the base, there is a current component for the base and equally for the emitter current that originates from the constant recombination-generation mechanism in the depletion region of the EB junction. This current, sometimes called the Space-Charge-Region (SCR) current, which depends on the width of the depletion region W_{dep} and the recombination-generation lifetime τ_{dep}. For the EB junction, a net generated current I_R (see Fig. A.1) can be expressed as [A.2]:

$$I_R = A_E q \frac{W_{dep} n_i}{2\tau_{dep}} \cdot exp(\frac{qV_{BE}}{2kT}) \,. \tag{A.13}$$

This current only affect the emitter current and does not affect the collector current, which in turn causes α_F to decrease. This effect is taken into account by defining the *recombination factor*, δ , as [A.2]:

$$\delta = \frac{I_{pE} + I_{nE}}{I_{pE} + I_{nE} + I_R} \approx \frac{I_{pE}}{I_{pE} + I_R} \,. \tag{A.14}$$

δ is essentially the ratio of the emitter current resulting from the injected minority carriers (i.e., $I_{pE} + I_{nE}$) to the total emitter current including I_R. Since the injected holes from emitter I_{pE} are usually much higher than I_{nE} in an efficient emitter, I_{nE} can be neglected in (A.14). I_{pE} and I_R are both functions of V_{BE} but with different exponential factors. By replacing (A.9) and (A.13) in (A.14), δ can be expressed approximately as:

$$\delta \approx \frac{1}{1 + \frac{W_B W_{dep} N_B}{\tau_{dep} D_B n_i} \cdot exp(\frac{-qV_{BE}}{2kT})} \,, \tag{A.15}$$

which remains a function of V_{BE}. When V_{BE} is adequately high, i.e., $V_{BE} \gg 2kT$, δ can be approximated as 1, which then does not affect α_F. However, when V_{BE} reduces and approaches 2kT or even lower, the effect of δ becomes more apparent and starts to affect α_F and eventually affects the ideal transistor behavior. This, as discussed in Sec. 3.3, is also the reason for the reduced β at low V_{BE} values.

Apart from the three components of the base current discussed above, there are two negligible current components originating from the CB junction. These currents are caused, respectively, by the holes injected into the base from the reverse biased CB junction and by the recombination-generation current in the CB junction. However, since the CB junction is reverse biased and has a large barrier potential, both of these

currents are negligible compared to the other three base current components.

A.3 Temperature dependence of I_S

I_S is a strong function of temperature due to the temperature dependence of parameters involved in (A.8): n_i^2, D_B, and W_B.

The dependence of n_i to temperature is illustrated in Fig. A.2 (a) for silicon, which can theoretically be expressed as:

$$n_i^2(T) = N_c N_v \cdot exp(\frac{-E_g}{kT}) \,, \tag{A.16}$$

where $E_g(T)$ is the energy gap between the conduction and valence band of silicon, and N_c and N_v are the effective density of states in the conduction and valence bands. Taking into account that $N_v = 2 \cdot (2m_p^* kT/h^2)^{1.5}$ and $N_c = 2 \cdot (2m_n^* kT/h^2)^{1.5}$, n_i^2 can be expressed as [A.1]:

$$n_i^2(T) = 2 \cdot (m_p^* m_n^*)^{1.5} \cdot \left(\frac{2\pi kT}{h^2}\right)^3 \cdot exp(\frac{-E_g(T)}{kT}) \,, \tag{A.17}$$

where m_n^* and m_p^* are, respectively, the equivalent masses of the electrons and holes, and h is the Plank constant. The energy gap of silicon $E_g(T) = E_c(T) - E_v(T)$ also varies with temperature as illustrated in Fig. A.2 (b). It is usually modelled as a non-linear function of temperature as [A.3]:

$$E_g(T) = E_{g0} - \frac{a_1 T^2}{T + a_2} \,, \tag{A.18}$$

where E_{g0} is the bandgap energy of silicon at 0 K, which is 1.15 eV; a_1 and a_2 are constants [A.3]. In the temperature range between -55 °C and 150 °C, however, $E_g(T)$ and, similarly, the bandgap voltage $V_g(T) = q \cdot E_g(T)$ can be approximated as a linear function of temperature as:

$$E_g(T) = q \cdot V_g(T) = q \cdot (V_{g0} - \alpha_g T) \,, \tag{A.19}$$

where V_{g0} is the extrapolated bandgap voltage of silicon at 0 K and α_g is its temperature dependence.

D_B in the base region is a function of hole's mobility μ_p. Since at equilibrium, diffusion and drift currents need to be equal according to the *Einstein equation*, D_B is then equal to $\mu_p(T) \cdot kT/q$. Apart from the presence of T in this equation, $\mu_p(T)$ is

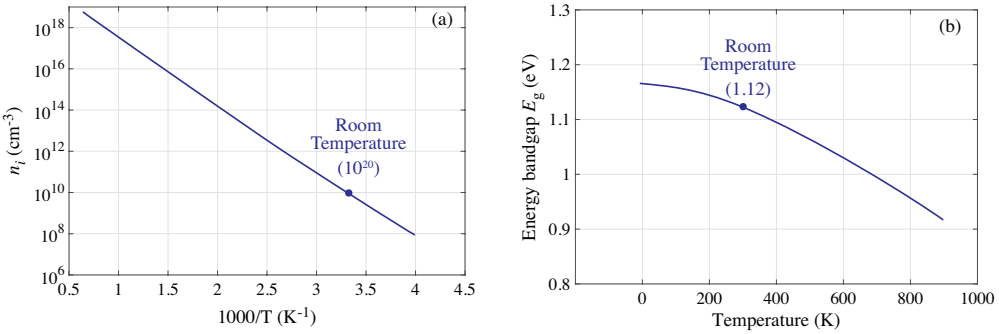


Figure A.2: Temperature dependece of: (a) intrinsic carrier concentration of silicon, n_i and (b) energy gap of silicon, E_g.

also heavily temperature-dependent. The temperature dependence of $\mu_p(T)$ is caused by two main mechanisms both of which relate to the scattering of carriers. The first mechanism, denoted as $\mu_l(T)$, is *phonon-scattering*, which dominates the overall scattering for $T >\sim -150$ °C and it is theoretically proportional to $T^{-1.5}$. Practical measurements, however, show the dependence of $T^{-2.2}$ to $T^{-2.4}$ [A.4]. For lower temperatures $T <\sim -150$ °C, the second scattering mechanism dominates, which is related to the *coulomb-scattering* of ionized dopant. This is denoted as $\mu_i(T)$, which is theoretically proportional to $T^{+1.5}$. The overall $\mu_p(T)$ is then has the combined effect of the two mechanisms with counter-dependence to temperature that can be expressed as [A.4]:

$$\mu(T) = \left(\frac{1}{\mu_l(T)} + \frac{1}{\mu_i(T)} \right)^{-1} . \tag{A.20}$$

$\mu_p(T)$ might have a peak at a certain temperature depending on the doping level and the counteraction of the two mechanisms in (A.20). A graphical view of $\mu(T)$ variations is shown in Fig. A.3 (a) for different doping concentrations and for both μ_p and μ_n. For low levels of doping concentration, the peak never happens, and the first scattering mechanism dominates in a wider temperature range. For vertical PNPs, since the base has low doping, $\mu(T) \sim \mu_l(T)$ can be expected for temeprtaures above -55 °C.

Besides the two scattering mechanisms, there are other second-order mechanisms that slightly affect the temperature-dependence of $\mu(T)$. As a result of this, in practice, the empirical model for mobility is found to be more useful, which can be ex-

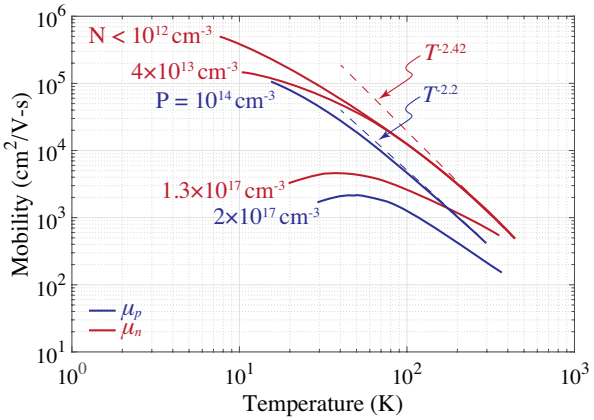


Figure A.3: Temperature dependence of mobility for different doping concentrations.

pressed as:

$$\mu(T) = \mu(T_r)\left(\frac{T}{T_r}\right)^{\alpha_{\mu p}} , \tag{A.21}$$

where $\mu(T_r)$ is the value of μ at an arbitrary temperature T_r and α_μ is a fitting parameter. Some works suggest more complicated models for α_μ to obtain a better fit with the empirical model to the measured μ, in which α_μ is also temperature-dependent [A.5].

W_B can also change with temeprature. Although the physical width of base is fixed, it changes as a result of change in the depletion region width in the two junctions. This effect is known as *early effect*, and will be further discussed in Sec. A.4.2.

Considering all the temperature dependences that are discussed in this section, and assuming a linear temperature dependence for $E_g(T)$, and a temperature-independent W_B, I_S can be approximated as:

$$I_S(T) = CT^\eta \cdot exp\left(\frac{-qV_{g0}}{kT}\right) , \tag{A.22}$$

where (η) is a constant that is related to the temerpature-dependence of mobility, and C is a fitting parameter.

A.4 Non-idealities of the $I_C - V_{BE}$ characteristic

A.4.1 High-level injection

The basic operation of a PNP explained in Sec. A.1 was based on the condition that the injected minority carriers from the emitter to the base p' are lower than the majority carrier in the base N_B. When $p' > N_B$, *high injection* occurs. The consequence of high injection is that hole and electron have a similar level of concentration in the base, i.e., $n \sim p$, allowing their values to be obtained from:

$$p \approx n \approx n_i \exp\left(\frac{qV_{BE}}{2kT}\right). \tag{A.23}$$

The collector current of the PNP in high-injection can be expressed as:

$$I_C \propto n_i \exp\left(\frac{qV_{BE}}{2kT}\right), \tag{A.24}$$

which indicates that the $I_C - V_{BE}$ curve in high injection has a lower exponent factor than the low injection condition. The current level at which the slope of $ln(I_C)$ versus V_{BE} reduces to half is called the high injection knee current I_{KF}. This slope reduction in $I_C - V_{BE}$ also causes β to drop in this region since I_B is not affected by this effect. Since (A.24) occurs at relatively high levels of current (e.g., $I_C > 100\ \mu$A), for the scope of this work, in which PNPs maximum bias current is $\sim 1\ \mu$A, it can be neglected. One way to shift I_{KF} to higher I_C levels is to increase N_B or decrease N_E. However, both options negatively affect the ideal operation of the transistor by reducing its emitter injection efficiency.

The high injection of minority carriers from the base into the emitter is almost impossible to occur since the doping level in the emitter N_E is high and making it difficult for the minority carriers from the base to reach the level of N_E.

A.4.2 Base-width modulation

The physical width of the base is defined by the fabrication process. The effective W_B, however, is affected by the voltage across the EB and CB junctions, which causes the width of their depletion regions to change.

As illustrated in Fig. A.4, W_B is affected by the change in the V_{BC}. This is known

as the *forward early effect*, which can be expressed as [A.6]:

$$W_B(V_{BC}) = W_B(V_{BC} = 0) \cdot \left(1 - \frac{V_{BC}}{V_A}\right) , \tag{A.25}$$

in which V_A is the early voltage measured from the $V_{CE} - I_C$ curve, having values of greater than 50 V in high performance BJTs. The change in W_B results in three main parameters of the BJT as described in the following:

(1) The saturation current I_S is inversely proportional to W_B as can be observed from (A.8). The change in I_S as a result of V_{BC} change can be expressed as:

$$I_S(V_{BC}) = I_S(V_{BC} = 0) \cdot \left(1 + \frac{V_{BC}}{V_A}\right) . \tag{A.26}$$

(2) The current gain β is also dependent on W_B and thus on V_{BC}. Due to fewer recombination chances for injected carriers in a shorter base, β increases with a decrease in V_{BC}. The dependence of β on the variation in W_B can be expressed as:

$$\beta(V_{BC}) = \beta(V_{BC} = 0) \cdot \left(1 + \frac{V_{BC}}{V_A}\right) . \tag{A.27}$$

(3) Base transit time τ_B can change when W_B changes. τ_B is proportional to W_B^2. This is a dynamic parameter of the device that defines the maximum operating frequency of the BJT, and is of less concern in the context of its DC characteristic.

The amount of depletion region width that falls on the base or collector depends on their relative doping concentrations. In some BiCMOS technologies, it is important to keep β independent from the electrical condition. In these cases, to keep the W_B less dependent on V_{BC}, a small region of the collector near the base-collector junction is made with a lower doping concentration compared to that of the base. This can absorb the depletion region width into the collector, thus reducing the variation in W_B. In order to avoid a high R_C, the rest of the collector is then made with substantially higher doping. In vertical PNPs, however, the collector is made from the p-substrate, and so there is little to no control over the collector doping. It can be assumed that the base and collector doping are at similar levels, since they are both made with a low-doped well. As a result, the depletion region will be divided nearly equally between the collector and the base, thus a high W_B modulation, or equivalently low V_A is expected.

Similar to the forward early effect, W_B is also modulated as a result of change in

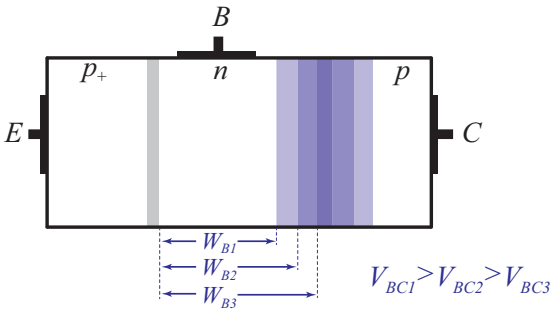


Figure A.4: Illustration of the forward early effect. The effective W_B changes due to variation in the CB depletion region width.

the EB depletion region when V_{BE} varies. This effect is known as the *reverse early effect*. The depletion region in the forward biased EB junction is substantially smaller compared to the reverse biased CB junction. On the other hand, since injection efficiency in the emitter requires N_B to be lower than N_E, the EB depletion region width predominantly falls in the base and thus change W_B. The change in I_S variation as a result of this W_B variation can be written as:

$$I_S(V_{BE}) = I_S(V_{BE} = 0) \cdot \left(1 + \frac{V_{BE}}{V_{AR}}\right) , \tag{A.28}$$

where $V_{AR}(\neq V_A)$ is the reverse early voltage. By applying (A.28) to the ideal $I_C - V_{BE}$ Equation of (A.7) and rearranging it, V_{BE} can then be expressed as:

$$\begin{aligned} V_{BE} &\approx \frac{1}{1 - \dfrac{kT}{qV_{AR}}} \cdot \frac{kT}{q} ln\left(\frac{I_C}{I_S}\right) \\ &\approx n_F \frac{kT}{q} ln\left(\frac{I_C}{I_S}\right) , \end{aligned} \tag{A.29}$$

where n_F is the non-ideality factor of the BJT. Although n_F in (A.29) is a temperature-dependent parameter, many circuit simulator define n_F as a fixed value that is slightly larger than 1 (e.g., 1.005) [A.7].

A.5 Non-linearity of the $V_{BE}(T)$ curve

The non-linearity of the V_{BE} causes curvature in bandgap voltage references and also non-linearity in temperature sensors. The non-linearity of V_{BE} can be observed by replacing the temperature dependence of I_S from (A.22) in (A.7) resulting in:

$$I_C(T) = CT^{\eta} \cdot exp\left(\frac{q(V_{BE}(T) - V_{g0})}{kT}\right). \tag{A.30}$$

Equivalently, V_{BE} can be expressed as:

$$V_{BE}(T) = V_{g0} + \frac{kT}{q} ln(I_C(T)) - \eta \frac{kT}{q} ln(CT). \tag{A.31}$$

This equation was derived from the physical properties of BJT devices and so, in principle, it should fit into the measured results when the constants are calculated properly. V_{g0} can be replaced with its value from the bandgap of silicon, η can be replaced with its value based on the temperature dependence of n_i and μ_p, and C can be replaced with its value based on A_E, W_B, the temperature dependence of V_g, and the proportionality factor of N_B. It has been shown that by using such physics-driven parameters, Equation (A.31) does not perfectly fit into the measured results of V_{BE} [A.8],[A.9]. According to [A.10], this discrepancy originates from the inaccurate model of the temperature charactersitic of $V_g(T)$, which was assumed to linearly change with T. To solve this, $V_g(T)$ can be replaced with a function that models its behavior more accurately as [A.3]:

$$V_g(T) = V_{g0} - \frac{a_1 T^2}{T + a_2}, \tag{A.32}$$

where a_1 and a_2 are constants. A second-order polynomial for $V_g(T)$ has also been suggested, in which $V_g(T)$ is expressed as [A.10]:

$$V_g(T) = a_1 - a_2 T - a_3 T^2, \tag{A.33}$$

where $a_1 : a_3$ are constants, and a_1 is not necessarily equal to V_{g0}. Applying these models has resulted in a reduction in the discrepancy between the physics-driven model in (A.31) and the measured results of V_{BE} with one order of magnitude, resulting in V_{BE} errors of less than 200 μV.

With a different approach, despite the discrepancy in the measured results, it has been shown that (A.31) can still be used to accurately model V_{BE} under the assumption that $V_g(T)$ is a linear function of temperature. This, however, required

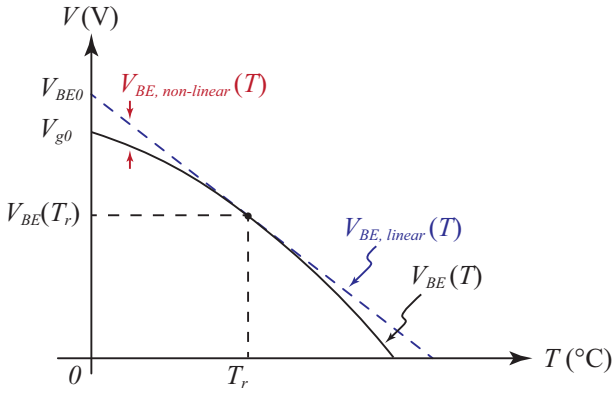


Figure A.5: Simplified illustration of V_{BE} non-linearity.

C, and V_{g0} to be replaced with their measured values from the $I_C - V_{BE}$ curve [A.11]. Using this method together with the empirically fitting parameters (A.33) obtained errors within 100 μV from the V_{BE} measurement. To obtain these fitting parameters, (A.33) needs to first be rearranged in a way so that the fitting parameters are found from voltage and current values as:

$$V_{BE}(T) = V_{g0}(1 - \frac{T}{T_r}) - \eta\frac{kT}{q}ln(\frac{T}{T_r}) + \frac{kT}{q}ln(\frac{I_C(T)}{I_C(T_r)}) + V_{g0}\frac{T}{T_r} + \frac{kT}{q}ln(T_r) + \eta\frac{kT}{q}ln(\frac{T_r}{C})\,, \tag{A.34}$$

where T_r is an arbitrary temperature at which the $I_C(T_r)$ and $V_{BE}(T_r)$ are found. The last three terms are all functions of T, which can then be lumped into $(T/T_r)V_{BE}(T_r)$, further simplifying (A.34) to:

$$V_{BE}(T) = V_{g0}(1 - \frac{T}{T_r}) - \eta\frac{kT}{q}ln(\frac{T}{T_r}) + \frac{kT}{q}ln(\frac{I_C(T)}{I_C(T_r)}) + \frac{T}{T_r}V_{BE}(T_r)\,. \tag{A.35}$$

Obtaining the exact behavior of V_{BE} requires knowing the bias current. The bias current can be replaced with $I_C(T) = I_C(T_r)(T/T_r)^m$, in which m determines its temperature dependence and $I_C(T_r)$ is its absolute value at T_r. For instance, $m = 0$ represents a constant current, whereas $m = 1$ represents a PTAT current. Replacing

$I_C(T)$ in (A.35) results in:

$$V_{BE}(T) = V_{g0}(1 - \frac{T}{T_r}) + (m - \eta)\frac{kT}{q} ln(\frac{T}{T_r}) + \frac{T}{T_r} V_{BE}(T_r)\,. \tag{A.36}$$

As illustrated in Fig. A.5, V_{BE} is usually considered as the sum of two functions: a linear function $V_{BE,linear}(T)$, which primarily determines V_{BE}, and as a non-linear function $V_{BE,nonlinear}(T)$. The linear function is a line that has a value of $V_{BE}(T_r)$ at T_r, and V_{BE0} at 0 K, and can be expressed as:

$$V_{BE,linear}(T) = V_{BE0}\left(1 - \frac{T}{T_r}\right) + \frac{T}{T_r} V_{BE}(T_r)\,. \tag{A.37}$$

To find V_{BE0}, the tangent of (A.36) at T_r needs to be found by taking its derivative at T_r and then finding the point where it crosses T = 0, which is:

$$V_{BE0} = V_{g0} + (\eta - m)\frac{kT_r}{q}\,. \tag{A.38}$$

The non-linear part of V_{BE} is then obtained by subtracting (A.37) from (A.36) resulting in:

$$V_{BE,non-linear}(T) = \frac{k}{q}(\eta - m)\left(T - T_r - Tln(\frac{T}{T_r})\right)\,. \tag{A.39}$$

One important observation from (A.39) is that V_{BE} non-linearity depends on m. Using m, which is larger than 1, can help to reduce the non-linearity [A.7]. Generating currents with $m > 1$, however, has its own limitations in terms of circuit design. Instead, combining two V_{BE} that have bias currents with $m = 0$ and $m = 1$ has been used to reduce the curvature of bandgap voltage references [A.12].

Another important observation from (A.39) is that V_{BE} non-linearity can be assumed to be a constant. When m is fixed, the non-linearity of V_{BE} is also fixed and does not spread with process. The assumption that η does not spread has also been made for the spread of I_S, in which the temperature dependence of I_S is considered to be process-independent.

A process-independent V_{BE} non-linearity has a great impact on designing bandgap voltage references and also temperature sensors. A process-independent V_{BE} non-linearity has been practically affirmed in many other works [A.13],[A.14] and is also used in Chapter 4 to obtain process-insensitive temperature sensor using the same average calibration parameters. In Chapter 5, the process-independent curvature of the bandgap voltage reference is used in *heater-assisted bandgap trimming*. When

V_{BE} non-linearity varies with process, neither of the low-cost calibration methods described in Chapter 4 and 5 are valid or they obtain lower accuracy.

A.6 Noise in the BJTs

A simple model for noise in a BJT is composed of three main sources: the shot noise of the collector, base currents, and the thermal noise of the base resistance. Other resistances, which are explained in the Sec. 3.4, also generate noise, but since they have lower values compared to the base resistance, they are negligible. The shot noise contribution from the base current has a power density of $I_{nB}^2 = 2qI_B$, and it is modelled as a current noise between the emitter and base terminals. Similarly, the shot noise of the collector current has a power density of $I_{nC}^2 = 2qI_C$, which is modelled as a current noise between the collector and the emitter terminals. When β is large (e.g., $\beta > 10$), I_{nB}^2 can be neglected compared to I_{nC}^2. Assuming that the PNP is diode-connected, this current will then flow through the equivalent small signal output impedance $\sim 1/g_m$ and create voltage noise with a power density of:

$$v_{n,V_{BE}}^2 = 2qI_C \cdot \left(\frac{1}{g_m}\right)^2 = \frac{2kT}{g_m}\,, \tag{A.40}$$

where g_m is the small signal transconductance of the PNP obtained from $g_m = qI_C/kT$. In this derivation, the noise of the current source that provides the bias current is neglected. When this bias current is also associated with the same level of noise to I_{nC}^2, the total output noise density in V_{BE} is simply doubled to $4kT/g_m$. The actual level of noise from the current source, however, depends on the bias circuit and the noise of the current mirrors.

Noise from the base resistance can be expressed as $V_{nR_B}^2 = 4kTR_B$. This noise is modelled as a voltage noise in series with V_{BE} and thus will be directly added to (A.40). When the bias current of the PNP is in the order of 1 μA or lower, the contribution of the base resistance is negligible compared to that of I_{nC}^2. Comparing these two requires $4kTR_B$ to be compared with $2kT/g_m$, or simply R_B to be compared with $1/2g_m$. Assuming $I_C = 1\mu$A, $1/(2g_m) = 12.5$ kΩ while R_B is usually in the order of few hundred ohms. When the bias current is lower than 1 μA, $1/(2g_m)$ increases even further making I_{nC}^2 even more dominant.

The noise in ΔV_{BE} can simply be obtained using (A.40) for two BJTs with col-

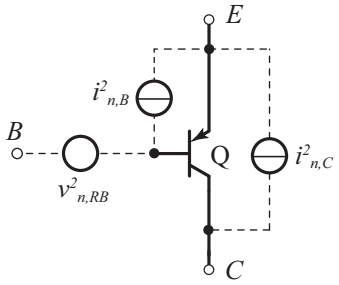


Figure A.6: Simple noise model for a BJT.

lector currents I_C and pI_C. As a result, this noise can be expressed as:

$$v^2_{n,\Delta V_{BE}} = v^2_{n,V_{BE}(I_C)} + v^2_{n,V_{BE}(p\cdot I_C)}$$
$$= \frac{2kT}{g_m} \cdot \left(1 + \frac{1}{p}\right) . \tag{A.41}$$

The low frequency noise behavior in a BJT deviates from the frequency-independent noise model provided in (A.40) and (A.41). To obtain a precise model for the low frequency noise, the mechanism of each current component in I_B, I_C and I_E needs to be considered. In most cases, however, a simple model has been found to be sufficient, in which only the low frequency noise associated with the base current is taken into account. Although the physical origin of low frequency noise in BJTs has not been conclusively determined, many studies relate it to surface states [A.15], while mobility fluctuations in the base is also suggested [A.16]. The low frequency noise of the base current can be modelled as [A.17]:

$$I^2_{n,B}(f) = \frac{KF \cdot I_B^{AF}}{f} , \tag{A.42}$$

where KF is a constant which is $\sim 2.15 \times 10^{15}$, and AF is a constant between 0.9 to 1.75 [A.17]. The corner frequency can be found, at which point the frequency-independent noise intersects with the frequency-dependent noise in (A.40). Assuming that I^2_{nC} is the dominant noise source for the frequency-independent noise, the corner frequency can be expressed as $f_c = KF/2qI_B^{(AF-1)}$. For typical values of KF and AF, and assuming $I_C = 1\mu$A, an f_c of less than 10 Hz will be obtained.

Bibliography

[A.1] C. Hu, *Modern Semiconductor Devices for Integrated Circuits*. Prentice Hall, 2010. [Online]. Available: https://books.google.nl/books?id=PosRbWdafnsC

[A.2] D. A. Neamen, *Semiconductor physics and devices: basic principles*, 3rd ed. New York: McGraw-Hill, 2003.

[A.3] Y. Varshni, "Temperature dependence of the energy gap in semiconductors," *Physica*, vol. 34, no. 1, pp. 149 – 154, 1967. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0031891467900626

[A.4] S. Sze and K. Ng, *Physics of Semiconductor Devices*. Wiley, 2006. [Online]. Available: https://books.google.nl/books?id=zHh-z5vZ7ZoC

[A.5] I. M. Filanovsky and A. Allam, "Mutual compensation of mobility and threshold voltage temperature effects with applications in CMOS circuits," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 7, pp. 876–884, July 2001.

[A.6] I. E. Getreu, *Modeling the Bipolar Transistor*. Elsevier, 1978.

[A.7] M. A. P. Pertijs and J. H. Huijsing, *Precision temperature sensors in CMOS technology*. Dordrecht, The Netherlands: Springer, Jan. 2006.

[A.8] A. Ohte and M. Yamagata, "A precision silicon transistor thermometer," *IEEE Transactions on Instrumentation and Measurement*, vol. 26, no. 4, pp. 335–341, Dec. 1977.

[A.9] K. E. Kuijk, "A precision reference voltage source," *IEEE Journal of Solid-State Circuits*, vol. 8, no. 3, pp. 222–226, June 1973.

[A.10] Y. P. Tsividis, "Accurate analysis of temperature effects in $I_C - V_{BE}$ characteristics with application to bandgap reference sources," *IEEE Journal of Solid-State Circuits*, vol. 15, no. 6, pp. 1076–1084, Dec. 1980.

[A.11] G. C. M. Meijer and K. Vingerling, "Measurement of the temperature dependence of the $I_C-(V_{BE})$ characteristics of integrated bipolar transistors," *IEEE Journal of Solid-State Circuits*, vol. 15, no. 2, pp. 237–240, Apr. 1980.

[A.12] P. Malcovati, F. Maloberti, C. Fiocchi, and M. Pruzzi, "Curvature-compensated BiCMOS bandgap with 1-V supply voltage," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 7, pp. 1076–1081, July 2001.

[A.13] S. Heidary Shalmany, D. Draxelmayr, and K. A. A. Makinwa, "A ±5 A integrated current-sensing system with ±0.3% gain error and 16 μA offset from −55 °C to +85 °C," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 4, pp. 800–808, Apr. 2016.

[A.14] S. Heidary Shalmany, D. Draxelmayr, and K. A. A. Makinwa, "A ±36 A integrated current-sensing system with a 0.3% gain error and a 400 μA offset from −55 °C to +85 °C," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 4, pp. 1034–1043, Apr. 2017.

[A.15] P. Lu, "Low-frequency noise in self-aligned bipolar transistors," *Journal of Applied Physics*, vol. 62, no. 4, pp. 1335–1339, Aug. 1987. [Online]. Available: https://doi.org/10.1063/1.339660

[A.16] T. G. M. Kleinpenning, "Location of low-frequency noise sources in submicrometer bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 39, no. 6, pp. 1501–1506, June 1992.

[A.17] J. C. Costa, D. Ngo, R. Jackson, N. Camilleri, and J. Jaffee, "Extracting $1/f$ noise coefficients for BJT's," *IEEE Transactions on Electron Devices*, vol. 41, no. 11, pp. 1992–1999, Nov. 1994.

Summary

This thesis explores the realization of a precision temperature sensor in CMOS technology that can be produced at low costs. At the start of this work, A BJT-based sensor in the target 0.16-μm technology could obtain inaccuracy of $\pm 0.15^{\circ}$C (3σ) in the range from -55°C to 125°C. This, however, shifted between two batches without its root cause being determined. Due to different mechanical stress, the result of this sensor also shifted from ceramic to plastic packages. Both of these shifts could only be corrected at the expense of more calibration, i.e., separate calibration for each batch, and for each package type. The need to re-calibrate separate batches and differently packaged samples made it impractical to use such a sensor as a low-cost product.

In this thesis, a sensor with higher accuracy is realized that achieves 3σ error of $\pm 0.06^{\circ}$C in the same range, while it does not require extra calibrations, and so can be produced with lower costs. In addition, two calibration methods are proposed based on using an on-chip heater that can help reduce the production costs of the sensor. Finally, the circuitry of the sensor is reused to accommodate extra functionality, and by doing so making more cost-effective use of its area.

Chapter 1 gives an introduction. It starts with a short history behind the use of BJTs for temperature sensing, and the fact that they are still popular in CMOS technology. It also gives a brief cost breakdown for the production of these sensors. It then discusses, in detail, why the state-of-the-art sensor in the target CMOS technology could not be produced at low costs. Methods used in this thesis to overcome these shortcomings are then explained.

Chapter 2 takes a step back and gives an overview of different methods for realizing temperature sensors in CMOS technology. It discusses four different types of sensor based on using resistors, MOSFETs, thermal-diffusivity, and BJTs. For each type of sensor, the operating principle, examples of readout circuits, reported accuracy based on the available literature, and calibration requirements are discussed.

Chapter 3 then focuses on BJT-based sensors, as this is the chosen method in this thesis. It starts with simple equations that govern an ideal BJT operation. The underlying physics behind the saturation current of BJTs, as the main determining

factor of their temperature dependence and source of spread, is discussed further. Non-ideal parameters of a BJT, such as limited β and series resistance are also discussed since they can cause additional errors to the sensor. Finally, the effects of mechanical stress caused by different packaging on the BJTs are discussed.

Chapter 4 discusses the first implementation in this thesis, which is about an accurate and process-insensitive sensor. It starts with an analysis of the error sources of the sensor's front-end. A proposed front-end that can deal with these errors and the sensor's ADC is then discussed. Measurements of this sensor over three different batches show that it obtains state-of-the-art accuracy, while eliminates the need for extra batch-calibration. It has been shown that the sensor's accuracy can be reproduced by a one-time calibration to extract a set of global parameters together with a single-point room temperature trim.

Chapter 5 discusses the second implementation in this thesis, which involves the development of calibration and trimming methods based on using an on-chip heater. These methods can help reduce the costs of producing the sensor. The first method, heater-assisted voltage-calibration, provides a fast two-point calibration without the need for external infrastructures except for an accurate external voltage reference. It has been shown that this calibration method can enhance the sensor's accuracy and also correct shifts of plastic packaging. The second method, heater-assisted bandgap trimming, provides an extremely low-cost trimming method that can correct BJT's spread without the need for accurate external references. Although the accuracy after this trimming is not as accurate as the first calibration method, it is still beneficial for noisy test environments, where accurate references are not available. The effectiveness of these two methods on the sensor are verified over a wide temperature range, and also with different packages.

Chapter 6 discusses the third implementation in this thesis, which aimed at making cost-effective use of the sensor's area. With a small overhead, the circuitry of the temperature sensor is re-used to also sense external voltages and capacitances. The result is a multi-function sensor that can sense temperature, while it can also digitize external voltage and capacitance.

Chapter 7 concludes this thesis by listing the main findings and suggesting three ideas for future work. Appendix A provides further reading about BJT devices.

Samenvatting

Dit proefschrift onderzoekt de realisatie van een nauwkeurige temperatuursensor in CMOS-technologie die tegen lage kosten kan worden geproduceerd. Bij aanvang van deze studie kon een temperatuursensor die BJT gebruikt in de 0.16-μm technologie een nauwkeurigheid behalen aan $\pm 0.15°\text{C}$ (3σ) in het bereik van $-55°\text{C}$ tot $125°\text{C}$. Dit verschoof echter tussen twee batches zonder dat de oorzaak bekend was. Het resultaat van de sensor verschoof ook van keramische naar plastic verpakkingen vanwege verschillende mechanische stress. Beide verschuivingen konden enkel worden gecorrigeerd ten koste van meer kalibratie, d.w.z. afzonderlijke kalibratie voor elke batch en voor elk pakkettype. De noodzaak om sensoren van afzonderlijke *batches* en verschillende verpakkingen opnieuw te kalibreren, maakte het onpraktisch om die sensor te gebruiken als lage kosten product.

In dit proefschrift wordt een sensor met een hogere nauwkeurigheid gerealiseerd die in hetzelfde bereik een nauwkeurigheid van $\pm 0.06°\text{C}$ (3σ) behaalt, en tegelijkertijd geen extra kalibraties vereist, en daarom met lagere kosten kan geproduceerd worden. Daarnaast worden twee kalibratiemethoden voorgesteld, gebaseerd op het gebruik van een warmtebron op een chip die meedraagt in de reductie van de productiekosten. Tot slot wordt het circuit van de sensor hergebruikt om extra functionaliteit te bieden. Hierdoor wordt op dit gebied ook kosten bespaard.

Hoofdstuk 1 bevat een inleiding. De inleiding geeft een kort overzicht van de geschiedenis achter het gebruik van BJT voor temperatuurdetectie en het feit dat deze nog steeds populair zijn in CMOS-technologie. Tevens geeft de inleiding een korte opsomming van de productiekosten van deze sensoren. Ook wordt besproken waarom de *state-of-the-art* sensor in de 0.16-μm CMOS-technologie in het verleden niet tegen lage kosten kon worden geproduceerd. Tot slot worden de methoden die in dit proefschrift worden gebruikt om deze tekortkomingen te verhelpen uitgelegd.

Hoofdstuk 2 geeft een overzicht van verschillende methoden die worden gebruikt bij het realiseren van temperatuursensoren in CMOS-technologie. Het gaat in op vier verschillende soorten sensoren op basis van het gebruik van weerstand, MOSFET, warmtediffusie en BJT. Voor elk type sensor wordt het volgende besproken: werkingsprincipe, voorbeelden van *readout circuits*, gerapporteerde nauwkeurigheid

op basis van de beschikbare literatuur en kalibratie eisen.

Hoofdstuk 3 richt zich op BJT-gebaseerde sensoren, aangezien dit de gekozen methode is in dit proefschrift. Het begint met eenvoudige vergelijkingen die een ideale BJT-bewerking bepalen. Tevens wordt de onderliggende fysica achter de *saturation-current* van BJT besproken, omdat het de belangrijkste bepalende factor van hun verspreiding is. Ook worden niet-ideale parameters van een BJT besproken, zoals de serie weerstand en β, aangezien deze extra fouten aan de sensor kunnen veroorzaken. Ten slotte worden de effecten van mechanische stress die veroorzaakt worden door verschillende verpakkingen op de BJT besproken.

Hoofdstuk 4 bespreekt de eerste implementatie, die gaat over een nauwkeurige en procesongevoelige sensor. Het begint met een analyse over de *error-sources* van de *front-end* van de sensor. Daarna wordt een voorgesteld front-end die met deze fouten kan omgaan besproken en de ADC van de sensor. Metingen over drie verschillende batches van deze sensor laten zien dat deze het meest nauwkeurige is, terwijl er geen extra batch-kalibratie nodig is. Het is aangetoond dat de nauwkeurigheid van de sensor kan worden gereproduceerd door een eenmalige kalibratie, samen met een één-punt kamertemperatuur trim.

Hoofdstuk 5 bespreekt de tweede implementatie, die de ontwikkeling van kalibratie en trimmethodes omvat op basis van het gebruik van een warmtebron op een chip. Deze methoden kunnen de productiekosten van de sensor reduceren. De eerste methode, *heater-assisted voltage calibration*, zorgt voor een snelle twee-punt kalibratie zonder dat externe infrastructuren nodig zijn, behalve een nauwkeurige externe voltage referentie. Het is aangetoond dat deze kalibratiemethode de nauwkeurigheid van de sensor kan verbeteren en daarnaast verschuivingen van plastic verpakkingen kan corrigeren. De tweede methode, *heater-assisted bandgap trim*, biedt een extreem goedkope trimmethode die de spreiding van BJT kan corrigeren zonder dat nauwkeurige externe referenties nodig zijn. Hoewel de nauwkeurigheid na dit trim niet zo nauwkeurig is als bij de eerste methode, is het toch gunstig voor testomgevingen met veel ruis, waar nauwkeurige referenties niet beschikbaar zijn. De effectiviteit van deze twee methoden op de sensor wordt geverifieerd over een breed temperatuurbereik, en ook met verschillende pakketten.

Hoofdstuk 6 bespreekt de derde implementatie in dit proefschrift, die gericht was op het kosteneffectief gebruiken van het sensorgebied. Het circuit van de temperatuursensor hergebruikt om ook externe voltages en capaciteiten te detecteren. Het resultaat is een multifunctionele sensor die de temperatuur kan meten, terwijl hij ook externe voltage en capaciteit kan digitaliseren.

Hoofdstuk 7 besluit dit proefschrift met een opsomming van de belangrijkste bevindingen en drie suggesties voor toekomstig werk. Bijlage A biedt meer informatie over BJT-apparaten.

Acknowledgements

This page is dedicated to those who have supported me throughout my PhD journey, have been parts of my life, or both, over the past decade.

Firstly, I express my gratitude to Kofi Makinwa, my supervisor and mentor, for granting me the opportunity to conduct research in his group and for guiding me through the process. He consistently encouraged me and kept an open door for discussions. Kofi, I admire your enthusiasm for the field, the positive energy you radiate, your active involvement, and the high standards you maintain in both work and writing. Thank you for all I have learned, which extends beyond research-related topics.

I extend my appreciation to Robert van Veldhoven, who supported this work at NXP Semiconductors. Without his support, several tape-outs and packaged devices crucial for this project would not have been completed on time. I am thankful to the managers at NXP, Hans Naus, and Jan van Geloven, for supporting this work and later giving me the opportunity to join their team. I also acknowledge the support of others at NXP, including Sha Xia, Berry Buter, and Ibrahim Candan.

I also like to thank Michiel Pertijs for our collaboration and for engaging in insightful technical discussions on various topics. I greatly admire his personality and his work on integrated temperature sensors. Additionally, I am grateful to Fabio Sebastiano for his motivational presence in the group and for sharing his valuable experiences. Fabio showed a great example of professionalism for me, which I admire.

I extend my thanks to the staff in the Electronic Instrumentation (EI) lab. Particularly Joyce and Karen, who have always been supportive. I am grateful to Zu-yao, Lukasz, and Ron for organizing labs and offering hands-on support.

I express my gratitude to colleagues and researchers in the EI lab for their company on many occasions, spending good times together, and sharing the PhD journey. Thank you to Jan, Ŭgur, Çağrı, Burak, Lorenzo, Pierluigi, Ehsan, Chao, Zhao, Fei, Junfeng, Shoubhik, Sining, Thajie, Hui, Long, Vincent, Shubham, Shardul, Said, and Rui. I appreciate the interactions with former EI researchers from whom I've gained valuable insights: Mahdi, Kamran, Zichao, and Quinwen. A special thanks to Saleh

Heidary Shalmany for his assistance in adapting to life in the Netherlands and for voluntarily supporting and sharing experiences in projects. I'd also like to thank Amir Ghaderi for his guidance in using COMSOL.

I had the chance to be part of Knowles Corporation in Switzerland over the past several years, where I wrote most of this thesis (on weekends). It was a rewarding experience being part of a great team and working alongside great colleagues.

I want to express my sincere gratitude to those with whom I've shared trust and friendship. In Delft, to Mohsen and Negin, Massoud and Mina, Morteza and Haleh, Milad and Bahar, Kianoush, and Davood. Beyond Delft, and in the Netherlands, to Shahryar and Mahlagha, Farshid and Pantea, with whom I could enjoy quality time and experience a sense of normalcy in life. To my friends in Germany who made me feel at home and let me experience the warmth of persian culture again: Sara, Sepideh, and Farzad. To Jeroen van Kleef, for the incredible experiences we shared, and the great time during our trip to Egypt. Special thanks to Hans Abel for showing me different dimensions of life and proving to me that intelligence and skills are not confined to academia or industry. Finally, I thank Ali, Mostafa, Amir, Saeid, and Keyvan for their lifelong friendship.

I wish to thank my parents, who have always believed in me even when I made mistakes. I love you both dearly, and I wish I could spend more time by your side. I am also thankful to Behrooz, my brother, who has had my back in every situation.

I am deeply grateful to my family in the Netherlands and all my in-laws, who have accepted me and treated me as one of their own. The unforgettable moments we shared, particularly with Lennert, Ralph, Olav and Iris, have given me hope to continue. A heartfelt thanks to my mother-in-law, Hannie, for her kindness and love. Your presence has shaped my perspective about life in the most positive way.

Finally, I am grateful to my dear Joëlle. Your love has left a mark on my heart. I am thankful for all the good moments we shared and the experiences that have contributed to our growth. I acknowledge that the past seven years have presented their challenges. I appologize for not devoting more attention to our relationship. I wish I could do better. You deserve nothing but the best, and my sincerest wish for you is a brighter future.

Bahman
Rotterdam, April 2024

List of publications

Journal papers

[1] **B. Yousefzadeh**, S. Heidary Shalmany and K. A. A. Makinwa, "A BJT-Based temperature-to-digital converter with ±60 mK (3σ) inaccuracy from −55°C to +125°C in 0.16-μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 4, pp. 1044−1052, April 2017.

[2] **B. Yousefzadeh** and K. A. A. Makinwa, "A BJT-Based temperature-to-digital converter with a ±0.25°C 3σ-inaccuracy from −40°C to +180°C using heater-assisted voltage calibration," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 2, pp. 369−377, Feb. 2020.

Conference papers

[1] **B. Yousefzadeh**, S. H. Shalmany and K. Makinwa, "A BJT-based temperature-to-digital converter with ±60 mK (3σ) inaccuracy from −70°C to 125°C in 160nm CMOS," in *Proc. IEEE Symposium on VLSI Circuits*, June 2016, pp. 192−193.

[2] **B. Yousefzadeh** and K. A. A. Makinwa, "A BJT-based temperature sensor with a packaging-robust inaccuracy of ±0.3°C (3σ) from −55°C to +125°C after heater-assisted voltage calibration," in *Proc. IEEE International Solid-State Circuits Conference (ISSCC)*, Feb. 2017, pp. 162−163.

[3] **B. Yousefzadeh**, W. Wu, B. Buter, K. Makinwa and M. Pertijs, "A compact sensor readout circuit with combined temperature, capacitance and voltage sensing functionality," in *Proc. IEEE Symposium on VLSI Circuits*, June 2017, pp. 78−79.

[4] **B. Yousefzadeh**, U. Sönmez, N. Mehta, J. Borremans, M. A. P. Pertijs and K. A. A. Makinwa, "A generic read-out circuit for resistive transducers," in *Proc. 6th International Workshop on Advances in Sensors and Interfaces (IWASI)*, June 2015, pp. 122−125.

Patents

[1] **B. Yousefzadeh**, Kamran souri, Kofi A. A. Makinwa, "Heater-assisted voltage calibration of digital temperature sensors, U.S. Patent 10605676, Mar, 31, 2020.

About the author

Bahman Yousefzadeh received the B.Sc. degree from Shahid Chamran University, Iran, in 2009, and the M.Sc. degree from Sharif University of Technology, Iran, in 2011 both in electrical engineering. He pursued his Ph.D. at Delft University of Technology, from 2014 to 2018, which was in collaboration with NXP Semiconductors. In 2019, he joined NXP Semiconductors, in Eindhoven, the Netherlands, and in 2020, he joined Knowles electronics, in Rapperswil, Switzerland.

He was a recipient of the Best Student Paper Award at the 2016 Symposium on VLSI Circuits and the 2016 Analog Devices Outstanding Student Designer Award.