A 130nW BJT-based Temperature Sensor with a Dual-Mode Front-End

By Jida Peng

Supervisor:

Prof. Dr. K.A.A. Makinwa

In partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering
At the Department of Microelectronics
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

To be defended on January 29th, 2024

Student number:	5501679
Thesis Committee:	Prof. Dr. K.A.A. Makinwa
	Dr. F. Sebastiano
	Dr. G. Wang

Abstract

In this thesis, a low-power, high-accuracy BJT-based temperature sensor is proposed, which consists of two main parts: a BJT-based dual-mode front-end (DMFE) and a tracking Delta-Sigma-Modulator (DSM). The front-end employs vertical PNP transistors, which, compared to NPNs, exhibit fewer non-idealities when biased at nA-level currents. The proposed DMFE generates accurate proportional-to-absolute-temperature (PTAT) and complementary-to-absolute-temperature (CTAT) voltages, while only consuming half the power of a conventional front-end. The temperature dependent ratio of these voltages is then digitized by an energy-efficient tracking DSM. The prototype sensor was fabricated in a 0.18µm CMOS process and has an active area of 0.228mm^2. Measurement results show that the sensor achieves an inaccuracy of ±0.15°C (3σ) over the industrial temperature range (-45°C to 85°C) after 1-point temperature calibration, which corresponds to a relative inaccuracy of 0.3%. It also achieves a 32mK resolution in 100ms, while consuming only 130nW. Its low power consumption, accuracy, and resolution make it suitable for IoT applications.

Acknowledgments

The past two years have made me grow both personally and professionally from a student to an engineer. Looking back, I would never have reached the end of this project without the knowledge and support I got. Here, I would like to express my great thanks to those people who supported and taught me during my whole journey.

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Kofi Makinwa, for all the opportunities, discussions, and guidance throughout the past year. His emphasis on the importance of designing simple and effective systems deeply influenced me when I tried to solve a problem with many choices. His trust in my abilities and the knowledge he shared with me encouraged me during the hardest time I went through. I deeply appreciated your patience and support these days that made me become a true engineer.

I owe a great deal to my daily supervisor, Nandor Toth, who provided me with a lot of help in both personal and professional life to make this project possible. Thanks to his true care and involvement in my project, I could get back on track when I was immersed in amounts of issues that should be solved.

I would also like to thank my digital support guy, Miloš Grubor, who spent many late nights helping me with Verilog coding and digital synthesizing even after he left the group. His interesting viewpoints on digital design always helped me during the boring digital layout time. I also want to thank a former post-doc, Dr. Teruki Someya, who worked on a similar project before. He helped me a lot in understanding the whole readout system and reviewed significant parts of my thesis even though he is now based in Japan.

I would like to extend my thanks to Angst+Pfister B.V., especially Dr. Guijie Wang, for sharing her industrial perspective with me and guiding me through the thesis. These discussions will undoubtedly influence my decision-making strategy in my career.

I also want to thank Dr. Fabio Sebastiano for taking the time and interest in my Thesis Committee. A special thanks to Dr. Michiel Pertijs for his comprehensive and intuitive PhD thesis.

A successful project needs much work in the background. People in the EI Lab offered great support in the past year. I will miss those discussions, food, and precious after-work time shared with them. I would like to thank the EI Lab technicians, Lukasz Pakula, Zu Yao Chang, Ron van Puffelen, for helping me with my measurement setup and PCB design, and for bonding lots of chips for me and guiding the soldering process. Many thanks also to the EI Lab researchers and PhD candidates, Sundeep, Karimeldeen, Zhong, Alireza, Eren, and many more, for offering their help in my thesis. I also want to thank my peers in this group, Sanja, Tanmeet, Shenyang, Avish, and Giorgos, for filling the office with energy and laughter every day.

It was my fortune to have a group of friends making a homelike life in the Netherland. Thank you, Weihong, Remy, Zitao, Shijie, Chenyao, Shuangmu, Zhiheng, Runsheng, Chenming, Keyu, Chieh, Arthur, and many more, for mindful, happy and relaxing experiences brought to me. I also want to thank other friends, Dongli, Dianlun, Jiayi, and Jinwei, for giving me their remote support from China.

Last but not least, I want to thank my family for their unqualified support in my study and life. Their love and full trust in me made me have courage to pursue whatever I wanted. Without these, I could never become the person who I am now.

Table of Contents

List of Figures

List of Tables

1 Introduction

Temperature sensors are widely used in diverse fields such as environmental monitoring, wearable biomedical devices, smart homes, and industrial Internet-of-Things (IoT) applications. These often involve portable, low-power devices with digital microprocessors, particularly in IoT applications. However, traditional temperature sensors, such as thermistors, typically feature a sizable discrete-component readout circuit with an analog output. An additional analog-to-digital converter (ADC) is then needed, which increases system size and complexity. Consequently, temperature sensors implemented in CMOS technologies have gained popularity due to their compact and lightweight design, low cost, and digital output.

The focus of this thesis is on the design of CMOS temperature sensors for IoT applications. Such sensors must prioritize low-power consumption to maximize battery life. In the case of a Lithium coin-cell battery, for example, a ten-year battery life translates to an average power dissipation of only 3µW [1]. Additionally, they must achieve a specified level of accuracy over a designated temperature range, for example, ±0.5°C inaccuracy from -40°C to 85°C in industrial applications and ±0.05°C inaccuracy from 35°C to 42°C for medical. Using the box method to normalize for the different temperature ranges [2], these specifications translate into relative inaccuracies of 0.8% and 1.4%, respectively. Low cost is also crucial in IoT applications, dictating a small die area and, at most, a 1-point room temperature trim.

1.1 State-of-the-art temperature sensors

In CMOS technology, low-power temperature sensors can be implemented in many ways. Based on a comprehensive survey of CMOS temperature sensors published over the last 25 years [3], Figure 1.1 compares the power consumption and relative inaccuracy (RIA) of various types of temperature sensors. Based on the requirements for IoT applications, i.e., sub-3µW power consumption, 1-point trim, and less than 0.8% RIA, BJT- and MOSFET- based temperature sensors are the best candidates.

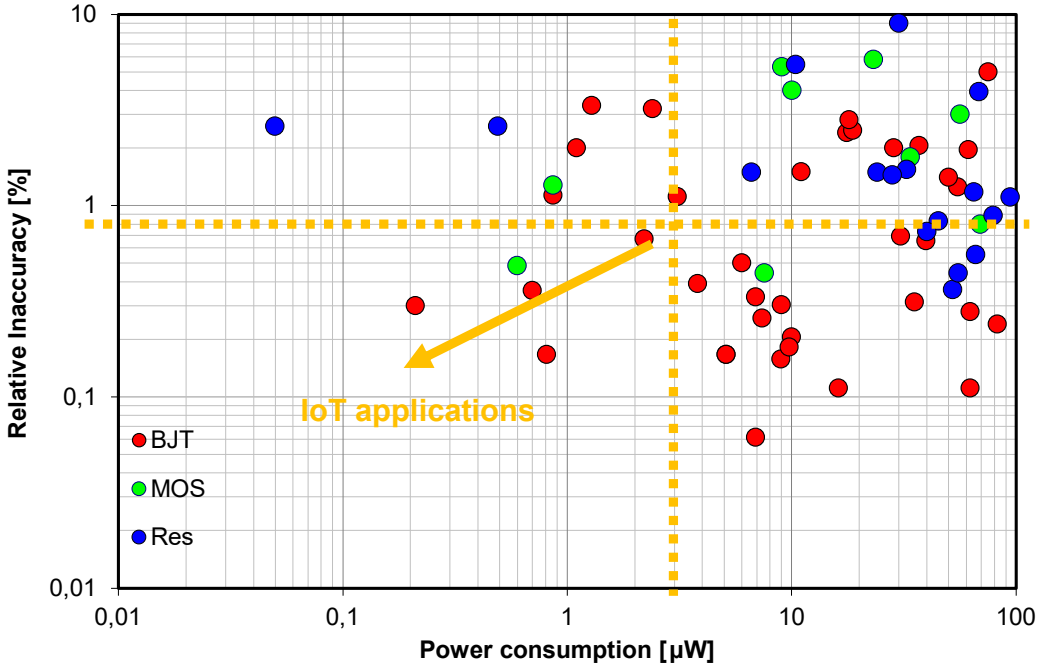


Figure 1.1 Relative Inaccuracy (less than 1-point calibration) versus Power consumption, plotted from [3].

1.2 Operating principle of BJT-based temperature sensors

The base-emitter voltage (V_{BE}) of a BJT is determined by its collector current (I_C). When I_C is significantly greater than the saturation current (I_S), V_{BE} can be expressed as follows:

$$V_{BE} \approx \frac{kT}{q} \ln\left(\frac{I_C}{I_S}\right), \tag{1.1}$$

where k is the Boltzmann constant (1.38×10^{-23}J/K), T is the absolute temperature in Kelvin degrees, and q is the electron charge (1.6×10^{-19}C). As detailed in [4], I_S can be expressed as:

$$I_S = \frac{kTAn_i^2\mu_p}{G_B}, \tag{1.2}$$

where A is the area of the emitter, G_B is the so-called Gummel number that expresses the number of impurities per unit area of the base, n_i is the intrinsic carrier concentration, and μ_p is the effective hole mobility. Due to the temperature dependence of μ_p and n_i^2, I_S is strongly temperature dependent ($\propto T^4$). As a result, V_{BE} is a complementary-to-absolute-temperature (CTAT) voltage with a slope of about -2.5mV/°C [4]. The effect of process spread on the I_C/I_S term causes V_{BE} spread, which is the main source of error in BJT-based sensors. In principle, this can be effectively corrected by trimming I_C at a single temperature [5].

As shown in Figure 1.2, BJT-based temperature sensors typically employ two BJTs biased at different current densities. In this case, the difference in their base-emitter voltages (ΔV_{BE}) will be proportional-to-absolute-temperature (PTAT) [6], which can be expressed as:

$$\Delta V_{BE} = V_{BE2} - V_{BE1} = \frac{kT}{q} \ln\left(\frac{pI_C}{I_S}\right) - \frac{kT}{q} \ln\left(\frac{I_C}{rI_S}\right) = \frac{kT}{q} \ln(pr), \tag{1.3}$$

where p is the ratio of their collector currents and r is the ratio of the emitter areas. The sensitivity of ΔV_{BE} is $k/q\cdot\ln(pr)$, which means that ΔV_{BE} is solely proportional to absolute temperature and is unaffected by the process spread of I_C/I_S. For $p\cdot r = 7$, the temperature sensitivity of ΔV_{BE} is 168μV/°C. This characteristic makes ΔV_{BE} a valuable and accurate temperature reference.

A temperature-independent reference voltage (V_{REF}) can be generated by adding a properly weighted ΔV_{BE} to V_{BE} [7] from the front-end, as shown in Figure 1.2. This allows accurate temperature measurements to be made without an external stable reference voltage source. The readout ADC combines V_{REF} and V_{PTAT} to generate a digital output (μ), which is a near-linear function of temperature:

$$\mu = \frac{V_{PTAT}}{V_{REF}} = \frac{\alpha \cdot \Delta V_{BE}}{\alpha \cdot \Delta V_{BE} + V_{BE}}, \tag{1.4}$$

where α represents the gain factor of $\partial V_{BE}/\partial \Delta V_{BE}$ [4].

With the help of two fitting parameters, A and B, the temperature in °C can then be expressed as:

$$T = A\mu + B, \tag{1.5}$$

where A is around 600, and B is around -273.

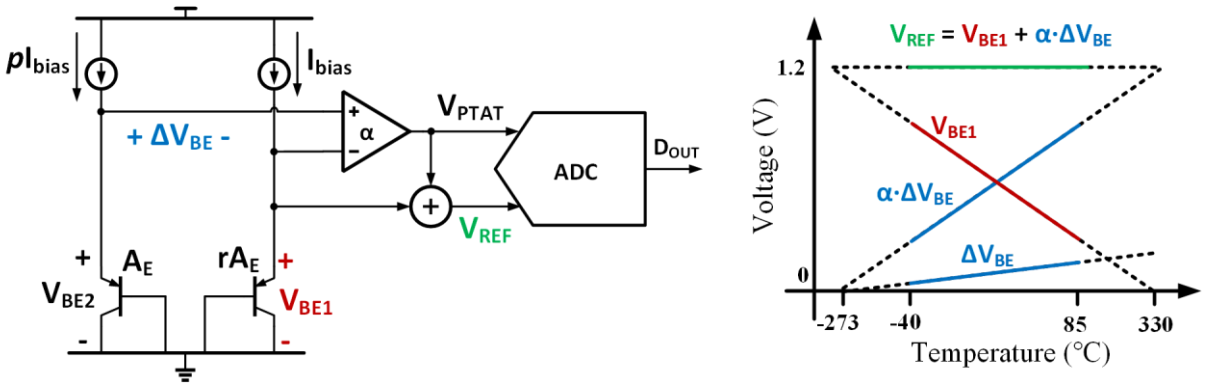


Figure 1.2 Sensing principle of BJT-based temperature sensors.

The front-end of a conventional BJT-based temperature sensor [8][9] is shown in Figure 1.3. To achieve high accuracy, it usually consists of two parts. One part is the bias circuit, which generates an accurate bias current (I_{bias}) by using an opamp to force ΔV_{BE} across a bias resistor (R_{bias}). The other is the bipolar core, which mirrors I_{bias} to two BJTs to generate supply-insensitive voltages V_{BE} and ΔV_{BE}. To compensate for the finite base current of the BJTs, a so-called beta-compensation resistor (R_{beta}) is used to modify I_{bias} such that the collector currents of the BJTs in the bipolar core are equal to $\Delta V_{BE}/R_{bias}$ [4].

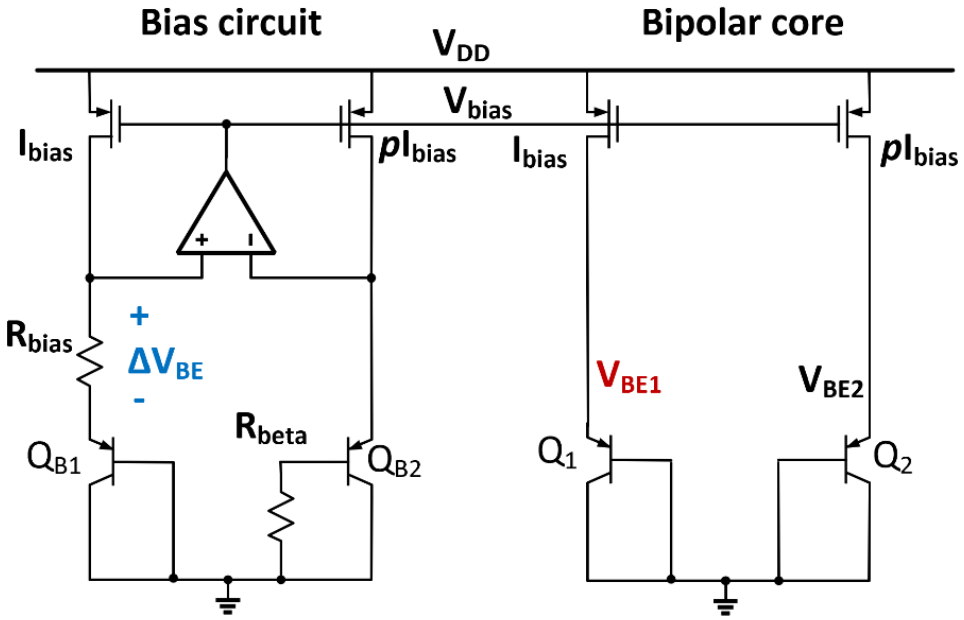


Figure 1.3 Conventional BJT-based front-end

In a comparable manner, MOSFET-based temperature sensors can also be realized, since the drain current (I_D) and the gate-source voltage (V_{GS}) of a sub-threshold-biased MOSFET also have an exponential relationship. However, both the oxide capacitance (C_{OX}) and the threshold voltage (V_{TH}) of MOSFETs are affected by process spread, which means that a 2-point trim is usually required to achieve less than 1% RIA [10].

1.3 Prior low-power temperature sensors

In a temperature sensor, the power consumption of its front-end is usually a significant part of its total power consumption. Therefore, the design of low-power front-ends will be discussed in the following:

In [11], a single-point trimmed 600nW all-CMOS temperature sensor is described, which uses dynamic threshold MOSTs (DT-MOSTs) as its temperature sensing elements. A DT-MOST is a standard MOSFET whose gate is connected to the bulk, which reduces the sensitivity of V_{GS} to the spread in gate-oxide thickness. As shown in Figure 1.4, A pair of DT-MOSTs with a 1:2 area ratio is biased by identical currents with a current-voltage mirror (CVM) to generate CTAT and PTAT voltages, V_{GS} and ΔV_{GS}, respectively. The value of V_{GS} (~ 0.3V) is only about half that of V_{BE} (~ 0.6V), allowing the sensor to operate from a sub-1V voltage supply. By applying dynamic-element-matching (DEM) and chopping to the CVM, the sensor achieves an inaccuracy of ±0.4℃ from -40 to 125℃, after a 1-point trim, while dissipating 600nW at room temperature.[3].

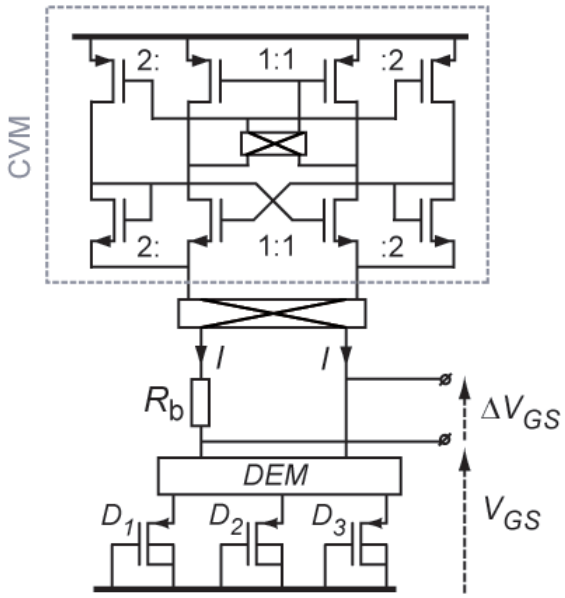


Figure 1.4 DTMOST-based sensor front-end [11].

A BJT-based temperature sensor capable of sub-1V operation is described in [12]. It avoids the headroom requirement of current sources by using capacitors to dynamically bias a pair of diode-connected BJTs. The basic operating principle is shown in Figure 1.5 for the case of a diode. A capacitor is first charged to the supply V_{DD} and then discharged across the diode. After an initial setting period, the voltage across the diode becomes independent of V_{DD} and is logarithmically related to the discharge time. By choosing two fixed discharge times, t_1 and t_2, two CTAT voltages, V_{D1} and V_{D2}, can be sampled on the capacitor, from which a PTAT signal ΔV_D ($\propto t_2/t_1$ = 32) can be generated. This sensor achieves an inaccuracy of ±0.15℃ from -55 to 125℃ with 1-point calibration while consuming 810nW from a 0.95V supply at room temperature.

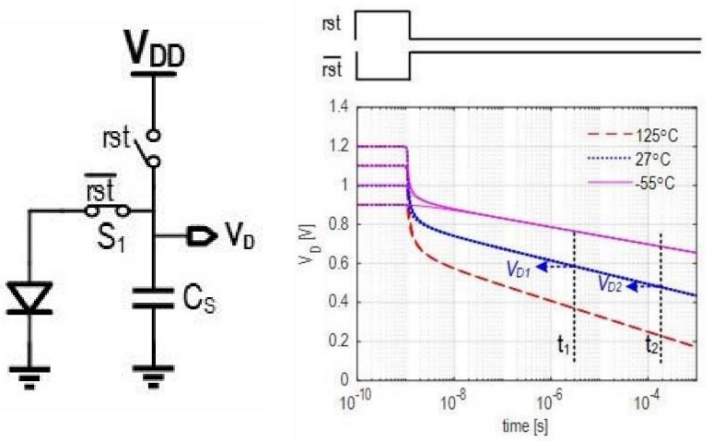


Figure 1.5 Front-end of the CB-BJT-based temperature sensor used in [12].

To reduce the power consumption further, a recently reported BJT-based temperature sensor employs a dual-mode front-end [13]. As shown in Figure 1.6, this design merges the bias circuit and BJT core of a conventional front-end [8][9] into a single dual-mode front-end (DMFE). In the pre-charge phase, the circuit is configured as a bias circuit to generate I_{BIAS}, and the corresponding gate voltage V_B of the PMOS current source is stored on the capacitor C_{bias}. This voltage is then held and used to generate the biasing currents of the reconfigured BJT core in the conversion phase. This modification approximately halves the front-end's power dissipation. The resulting sensor consumes 210nW power from a 1.25V supply while achieving a 1-point trimmed inaccuracy of ±0.15°C from -15°C to 85°C. However, its inaccuracy increases significantly above 85°C, as shown in Figure 1.7.

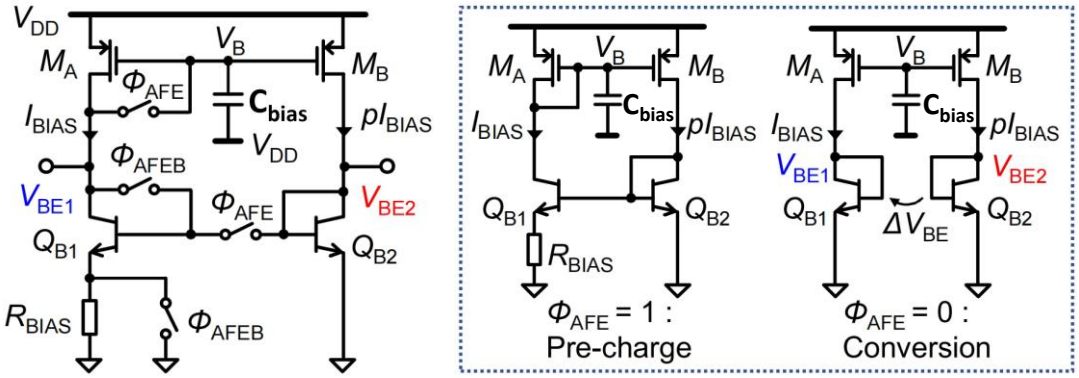


Figure 1.6 Proposed NPN-based DMFE [13].

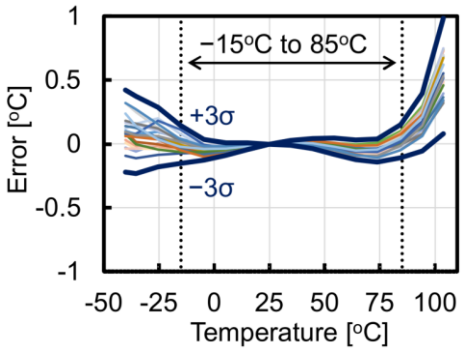


Figure 1.7 Measured inaccuracy of 20 samples with 1-point trim in [13].

The reason for this behavior was traced to the leakage current (I_{leak}) flowing through the (un-modelled) parasitic diode between the collectors of the sensing NPNs and the p-substrate, as shown in Figure 1.8. As temperature increases, I_{leak} increases exponentially, changing the current density ratio, and thus causing errors in ΔV_{BE}.

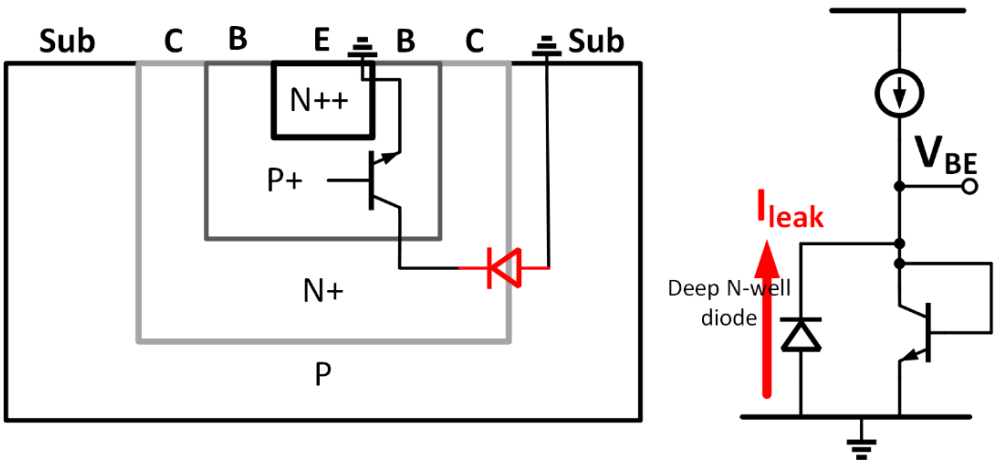


Figure 1.8 An NPN transistor model (left), an NPN core with an un-modeled diode (right).

This work aims to develop a sensor that consumes less power than the one in [13], while achieving good accuracy over a wider temperature range, thus making it more suitable for IoT applications. To do this, the NPNs in the DMFE are replaced by PNPs, which as shown in Figure 1.9, do not have parasitic diodes, and thus can be operated at lower bias currents and/or higher temperatures. Moreover, compared to the use of NPNs, the use of PNPs is more cost effective as they do not require an extra deep n-well implant.

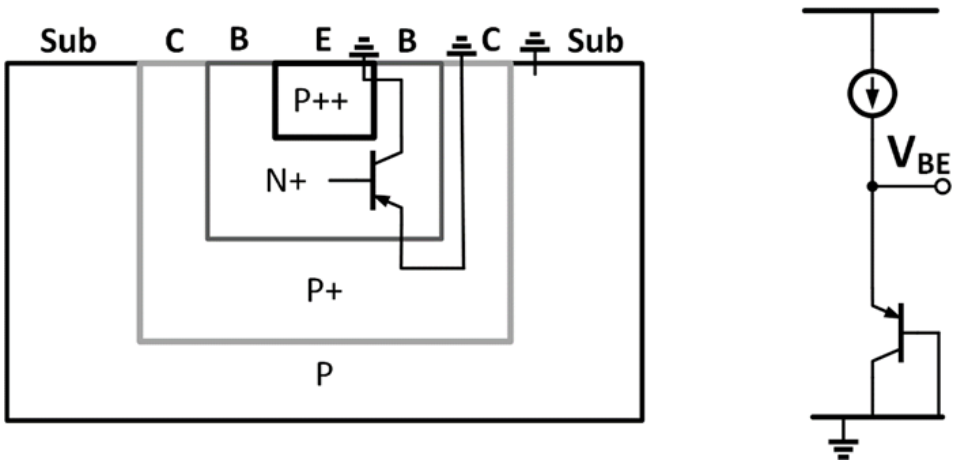


Figure 1.9 A PNP transistor model (left), a PNP core (right).

1.4 Target specification

The main goal of this work is to design a BJT-based temperature sensor with sub-100nW power consumption, but with an inaccuracy of about ±0.15 °C from -40 to 85°C after a 1-point trim, which translates into 0.24% RIA. As shown in Figure 1.10, achieving this target would represent a significant advance in the state of the art.

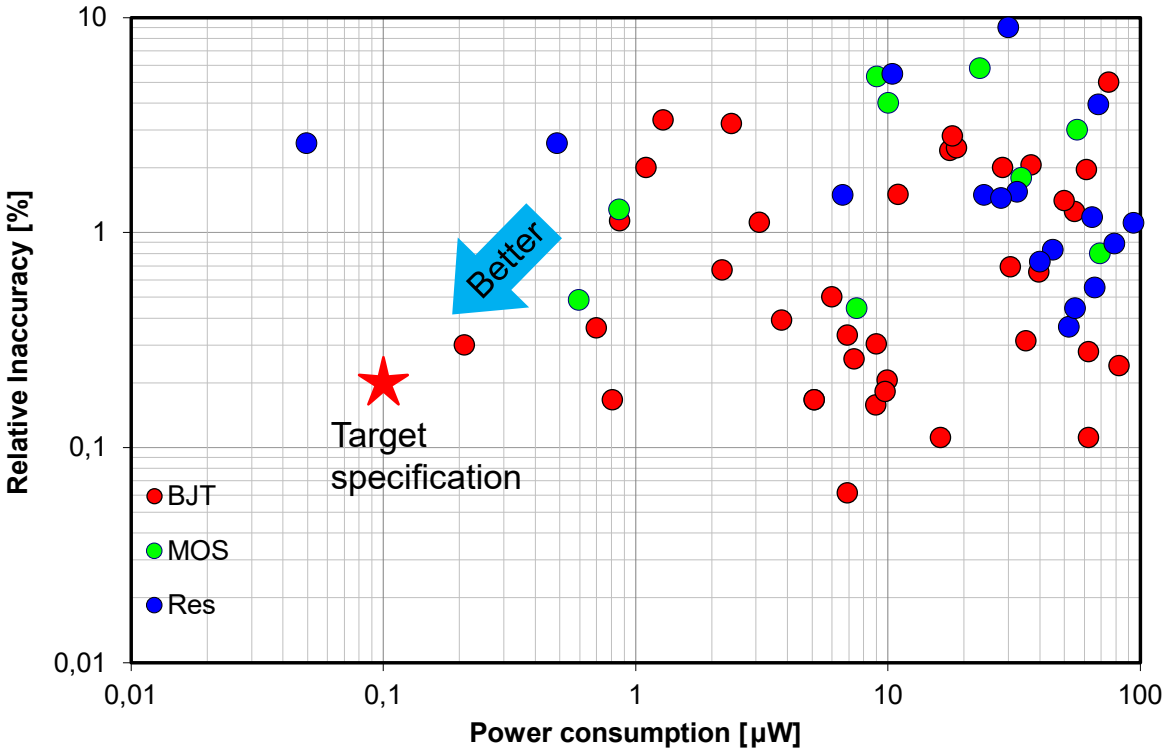


Figure 1.10 Target specification in the state-of-the-art low-power temperature sensors.

1.5 Organization of the thesis

The rest of the thesis is organized as follows. In Chapter 2, the proposed architecture and working principles are introduced. Also, the system-level requirements of the front-end and the readout circuit will be derived from the target design specifications. The detailed circuit-level implementations of the front-end, readout and the peripherals are described in Chapter 3. The sensor was taped out in a TSMC 180nm CMOS process, and the measurement results are presented in Chapter 4. They include accuracy, power, resolution, and power supply sensitivity results. Finally, Chapter 5 concludes with a discussion of the performance of this sensor and proposes possible future improvements.

2 Architecture considerations

Designing an accurate (≤0.24% RIA) temperature-to-digital converter (TDC) with low power consumption (≤100nW) is not a trivial task. System-level choices, e.g., about biasing currents and leakage tolerance, should be made before the circuit implementation. This chapter describes the architectural choices made in the design of both front-end and readout to meet the target accuracy while dissipating the lowest possible power.

2.1 Proposed PNP-based Dual-Mode Front-End (DMFE)

As in [13], the proposed PNP-based DMFE has two operating phases: a pre-charge phase (Φ_{AFE} = 1) and a conversion phase (Φ_{AFE} = 0), as shown in Figure 2.1. The PNP biased with pI_{bias} generates V_{BEH}, while the one biased with I_{bias} generates V_{BEL}. In the pre-charge phase, the front-end is configured as a bias circuit, which generates I_{bias} by using a bias amplifier to force ΔV_{BE} over R_{bias}. Its output (V_{bias}) is then sampled on a capacitor (C_{bias}). After V_{bias} settles the front-end enters the conversion phase, during which it is reconfigured as a BJT core that generates V_{BE} and ΔV_{BE}, while the bias amplifier is disabled. Compared with the front-ends of conventional BJT-based sensors [9], the bias amplifier only consumes power in the short pre-charge phase and is disabled in the conversion phase, resulting in significant power saving over the entire conversion period.

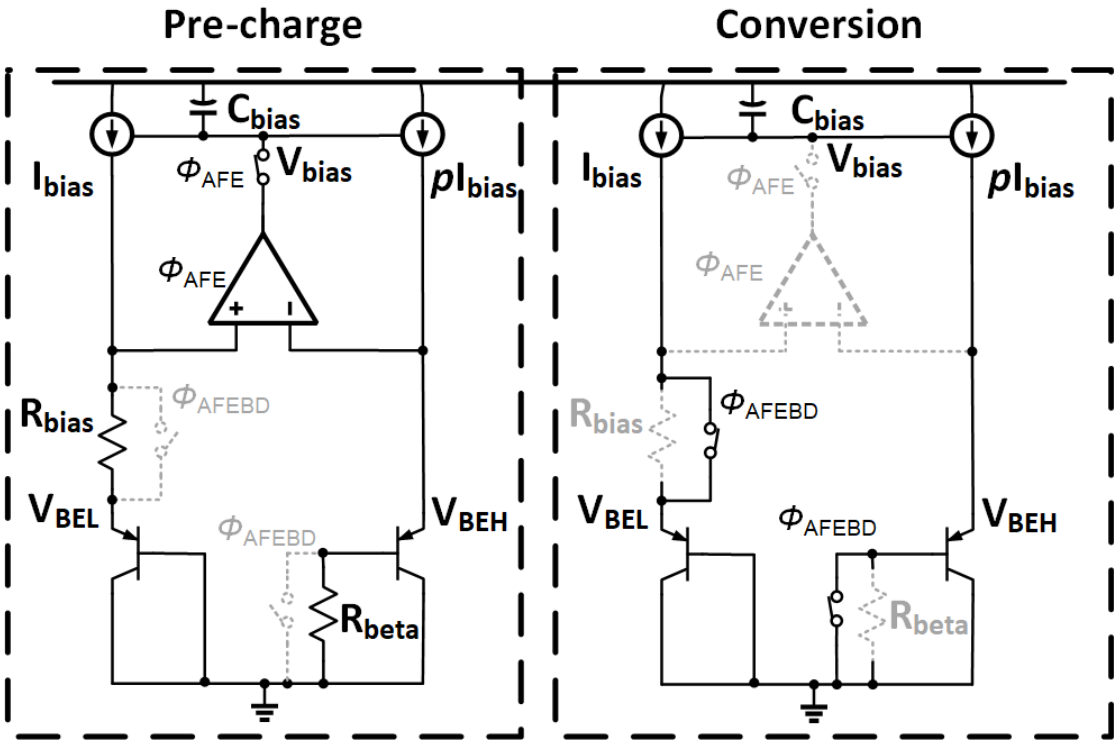


Figure 2.1 Two phases for the PNP-based DMFE.

2.1.1 Non-idealities in BJTs

As described in Chapter 1, the targeted inaccuracy of the proposed sensor is ±0.15°C from -40°C to 85°C. This is mainly determined by the non-idealities of the voltages generated by the front-end, V_{BE} and ΔV_{BE}. Thus, the temperature sensitivity of V_{BE} and ΔV_{BE}, which are calculated from Equations 2.1, 2.2, and 2.3 [4], can be used to make an error budget (<0.03°C) for each error source.

$$S_{V_{BE}}^{D_{OUT}}(T) = \frac{\partial D_{OUT}}{\partial V_{BE}} = \frac{A\partial\mu}{\partial V_{BE}} \approx -\frac{T}{V_{REF}}, \tag{2.1}$$

$$S_{\Delta V_{BE}}^{D_{OUT}} = \frac{\partial D_{OUT}}{\partial(\Delta V_{BE})} = \frac{A\partial\mu}{\partial(\Delta V_{BE})} \approx \frac{A-T}{V_{REF}}\alpha, \tag{2.2}$$

$$\alpha = -\frac{\partial V_{BE}}{\partial T} \Big/ \frac{\partial \Delta V_{BE}}{\partial T} = -\frac{\partial V_{BE}}{\partial T}\frac{q}{k\ln(pr)}, \tag{2.3}$$

Since the PNPs will be biased at nA current levels, I_S will become comparable to I_C, which adds more non-idealities to V_{BE} and ΔV_{BE}. The exact relationship between V_{BE}, ΔV_{BE}, and I_C is expressed as [4]:

$$V_{BE} = \frac{kT}{q}\ln\left(\frac{I_C}{I_S}+1\right), \tag{2.4}$$

$$\Delta V_{BE} = V_{BE1} - V_{BE2} = \frac{kT}{q}\ln\left(\frac{prI_C + I_S}{I_C + I_S}\right), \tag{2.5}$$

Since I_S is proportional to the emitter area (Equation 1.2) of the PNP, minimum-sized PNPs (2μm×2μm) were used to achieve the lowest I_C/I_S non-linearity. To set the current density ratio (CDR) of the two PNPs, either an area ratio (r) or a current ratio (p) can be used. As shown in Figure 2.2, for the same CDR and the same I_C/I_S non-linearity, emitter area scaling consumes more power. Thus, for low power, current scaling is the best choice. For high accuracy, p and, thus ΔV_{BE}, should be made as large as possible, however, this comes at the expense of power. As in [13], $p = 7$ was chosen as a good trade-off between power, accuracy, and implementation.

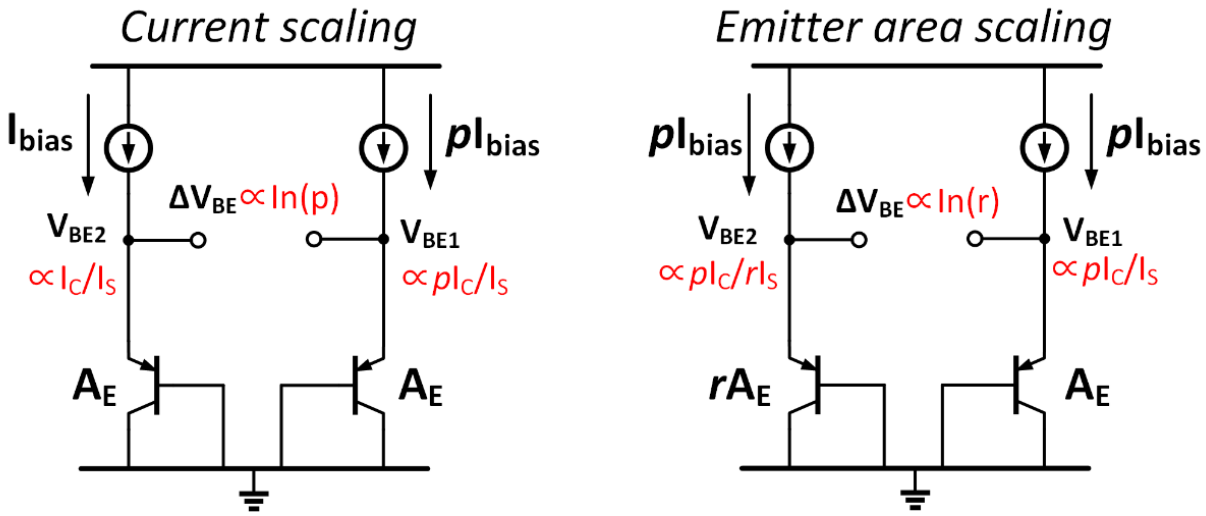


Figure 2.2 Two methods with the same CDR ($p = r$) and the same I_C/I_S non-linearity .

From [4], the effect of current ratio variation δp on ΔV_{BE} is given by:

$$\Delta V_{BE} = \frac{kT}{q}\ln(p+\delta p) \cong \Delta V_{BE,\delta p=0} + \frac{kT}{q}\frac{\delta p}{p}, \tag{2.6}$$

For $p = 7$, the temperature errors due to errors in ΔV_{BE} are shown in Figure 2.3 as a function of bias current and temperature. It can be seen that the effect of I_S non-linearity is negligible up to 85°C. However, at low temperatures, the PNP's limited base-collector current gain (β) introduces additional non-idealities in ΔV_{BE}, resulting in a large temperature error. Therefore, I_{bias} should be larger than 1nA to keep the absolute temperature error smaller than 0.03°C.

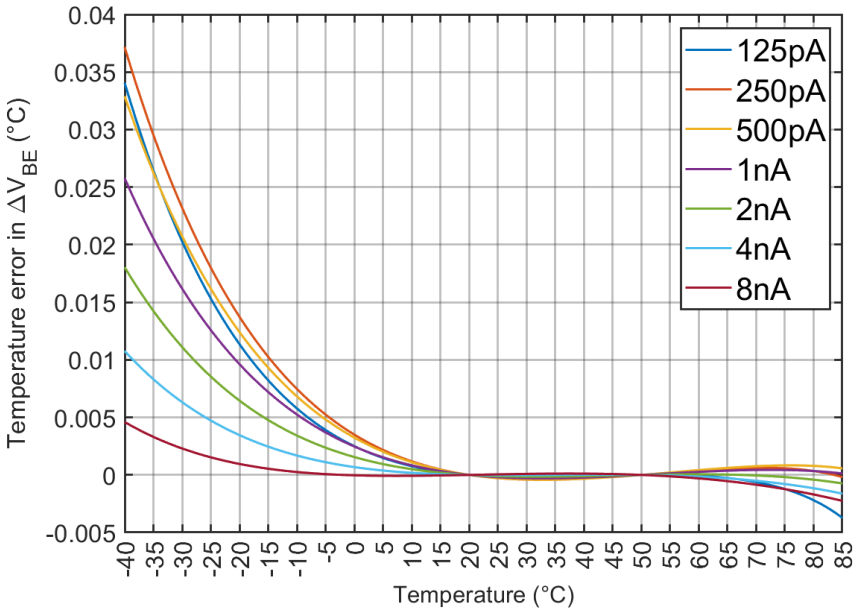


Figure 2.3 Temperature error introduced by ΔV_{BE} error from -40°C to 85°C over different bias currents.

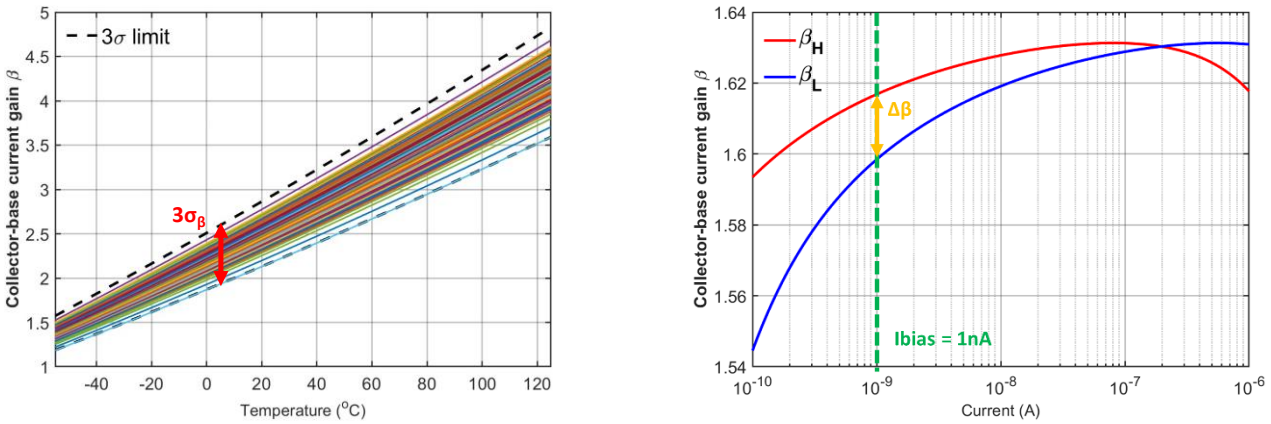
In a BJT, V_{BE} is determined by I_C. However, the PNPs used in this design can only be biased via their emitters. As a result, the base current (I_B) needs to be considered when determining V_{BE}, where β (= I_C/I_B) defines the loss caused by I_B. The exact relationship between V_{BE} and I_{bias} considering finite β is expressed as [4]:

$$V_{BE} = \frac{kT}{q}\ln\left(\frac{I_C}{I_S}\right) = \frac{kT}{q}\ln\left(\frac{I_E - I_B}{I_S}\right) = \frac{kT}{q}\ln\left(\frac{I_{bias}}{I_S}\frac{\beta}{\beta+1}\right), \tag{2.7}$$

The simulated β at different biasing currents is plotted in Figure 2.4 over temperature. The spread in β introduces an additional error in V_{BE}. The 1-σ spread in β is defined as σ_β. From Equation 2.7, the 3σ error in V_{BE} can be calculated as:

$$3\sigma_{V_{BE},\beta} = \frac{kT}{q}\ln\left(\frac{I_{bias}}{I_S}\cdot\frac{\beta+3\sigma_\beta}{\beta+3\sigma_\beta+1}\right) \cong V_{BE|3\sigma_\beta=0} + \frac{kT}{q}\cdot\frac{1}{1+\beta}\frac{3\sigma_\beta}{\beta}, \tag{2.8}$$

From Figure 2.4 (a), when biased at 1nA at low temperatures, β is close to 1, leading to significant error in V_{BE}. Specifically, β and σ_β vary with the temperature. From Equation 2.1 and Equation 2.8, the non-trimmable 3σ error of V_{BE} due to the variation in β and σ_β is ±0.05°C, as shown by the black curve in Figure 2.5.



(a) β spread from -55°C to 125°C for I_{bias} =1nA, (b) β spread with I_{bias} from 100pA to 1μA at -40°C.

Figure 2.4 Spread of β with 2μm×2μm PNPs.

As shown in Figure 2.4 (b), β_L is the β of the PNP biased with I_{bias} while β_H is the β of the PNP biased with $7I_{bias}$. When I_{bias} is 1nA, their difference $\Delta\beta$ (= β_H - β_L) is 0.018. Then ΔV_{BE} can be expressed as:

$$\Delta V_{BE} = V_{BEH} - V_{BEL} = \frac{kT}{q} \cdot \ln\left(p \cdot \frac{\beta_H}{\beta_H + 1} \cdot \frac{\beta_L + 1}{\beta_L}\right) = \frac{kT}{q}\ln(p) + \frac{kT}{q} \cdot \ln\left(\frac{\beta + \Delta\beta}{\beta + \Delta\beta + 1} \cdot \frac{\beta + 1}{\beta}\right), \quad (2.9)$$

When β is close to 1, the error in ΔV_{BE} becomes quite significant. Denoting the 1-σ spread in $\Delta\beta$ as $\sigma_{\Delta\beta}$, the resulting 3σ error is expressed as:

$$3\sigma_{\Delta V_{BE},\Delta\beta} = \frac{kT}{q}\ln(p) + \frac{kT}{q}\ln\left(\frac{\beta + 3\sigma_\beta + 3\sigma_{\Delta\beta} + \Delta\beta}{\beta + 1 + \Delta\beta + 3\sigma_\beta + 3\sigma_{\Delta\beta}} \cdot \frac{\beta + 3\sigma_\beta + 1}{\beta + 3\sigma_\beta} \cdot \frac{\beta_L}{\beta_L + 1} \cdot \frac{\beta_H + 1}{\beta_H}\right),$$
$$\cong \Delta V_{BE|3\sigma_\beta=0\ \&\ 3\sigma_{\Delta\beta}=0} + \frac{kT}{q} \cdot \frac{1}{\beta + 1 + \Delta\beta}\frac{3\sigma_{\Delta\beta}}{\beta}, \quad (2.10)$$

As will be discussed in the following section, the mismatch of $\Delta\beta$ can be mitigated by dynamic element matching (DEM) techniques. Thus, only the process spread in $\Delta\beta$ ($3\sigma_{\Delta\beta}$) is considered. Considering the temperature dependence of $\sigma_{\Delta\beta}$, $\Delta\beta$, and β, then by combining Equation 2.2 and Equation 2.10, the 3σ error in ΔV_{BE} is ±0.08°C, as shown by the red curve in Figure 2.5. The combined error of β and $\Delta\beta$ spreads about ±0.13°C, which exceeds the error budget, as shown by the blue curve in Figure 2.5.

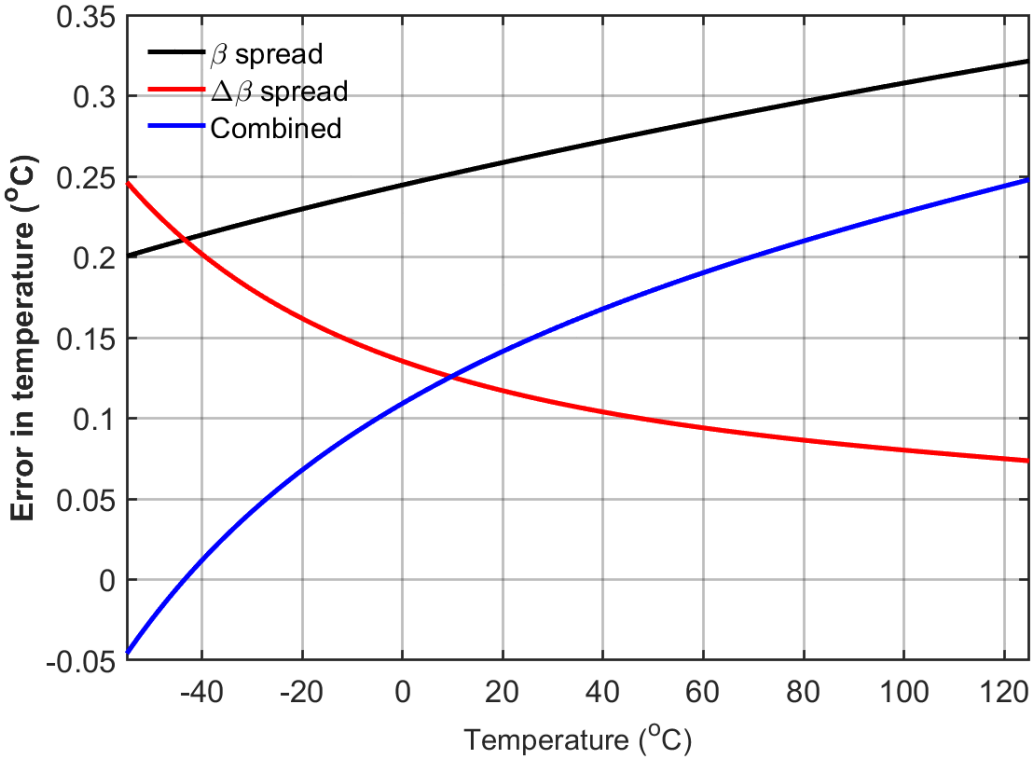


Figure 2.5 Temperature error caused by spread in β and $\Delta\beta$

The technique known as β-compensation mitigates this problem by modifying the PTAT bias circuit to generate a β-dependent I_{bias}, as shown in Figure 2.6. This is done by adding a β-compensation resistor (R_{beta}) of $R_{bias}/7$ in series with the base of Q_2. The op-amp in the feedback loop then ensures that [4]:

$$I_{bias} = \ln(p)\frac{kT}{q}\frac{1}{R_{bias}}\frac{1+\beta}{\beta}, \quad (2.11)$$

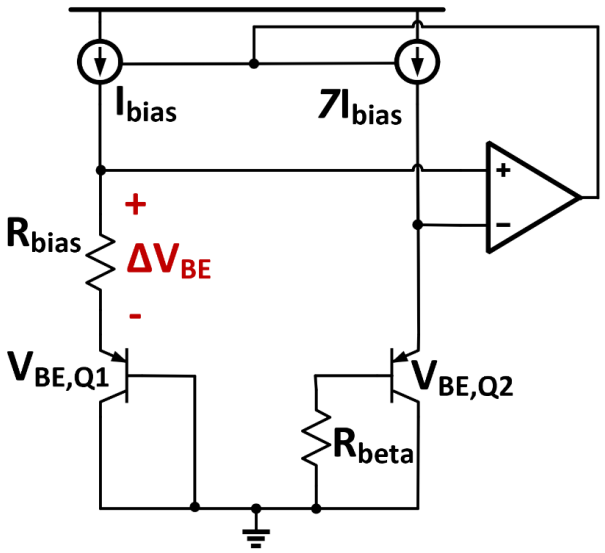


Figure 2.6 Modified PTAT bias circuit with β-compensation [4].

The new relationship between ΔV_{BE} and V_{BE} is then given by:

$$V_{BE} = \frac{kT}{q}\ln\left(\frac{I_{bias}}{I_S}\frac{\beta}{\beta+1}\right) = \frac{kT}{q}\ln\left(\frac{\Delta V_{BE}}{R_{bias}I_S}\right), \tag{2.12}$$

where V_{BE} is independent of β spread because the PNP's collector current is now $\Delta V_{BE}/R_{bias}$. However, beta compensation only cancels the influence of β in V_{BE}, but not the temperature error in ΔV_{BE}. However, the residual error after β-compensation is only ±0.08℃, which is acceptable for a ±0.15℃ inaccuracy requirement.

2.1.2 Bias amplifier

As shown in Figure 2.1, a bias amplifier is used in the DMFE to generate I_{bias} (=1nA) by forcing an accurate ΔV_{BE} over R_{bias}. I_{bias} is calculated as shown in Equation 2.13. Thus, R_{bias} is 50MΩ.

$$I_{bias} = \frac{\Delta V_{BE}}{R_{bias}} = \frac{\frac{kT}{q}\ln(p)}{R_{bias}}, \tag{2.13}$$

Due to the bias amplifier's finite open-loop gain ($A_{OL,op\text{-}amp}$) and mismatch, it will have a finite overdrive voltage. As discussed in [4], this voltage is effectively added to the ΔV_{BE} forced across R_{bias}, leading to errors in I_{bias}, and, in turn, to errors in the V_{BE} generated by the BJT core. In this design, the amplifier's offset, which would otherwise cause a large (±2℃) error, is mitigated by chopping over two conversions. However, the overdrive voltage caused by finite $A_{OL,op\text{-}amp}$ cannot be chopped and will affect V_{BE} in the conversion phase. The resulting V_{BE} error ($V_{BE,error,bias}$) is expressed as [4]:

$$A_{CL} = \frac{\Delta V_{out}}{\Delta V_{in}} = A_{OL,opamp}g_{m,CM}R_{bias}, \tag{2.14}$$

$$V_{BE,error,bias} = \frac{\Delta V_{BE}}{A_{CL}\cdot\ln(p)} < \frac{\Delta T_{err}}{S_{V_{BE}}^{D_{out}}} = \frac{0.03℃ \times V_{ref}}{300} = 120\mu V, \tag{2.15}$$

where A_{CL} is the close-loop gain, ΔV_{IN} is the overdrive voltage present at the input of the bias amplifier, ΔV_{OUT} is the voltage error added on ΔV_{BE}, and $g_{m,CM}$ is the transconductance of the current mirrors. In this case, $g_{m,CM}R_{bias}$

can be expressed as $g_{m,CM}/I_{bias}\cdot\Delta V_{BE}$, which is only dependent on the working region of the current mirrors. When they are biased in the strong inversion region ($g_m/I_d = 10$), $g_{m,CM}R_{bias}$ is $10\Delta V_{BE}$. Consequently, the open-loop gain of the bias amplifier must be larger than 53dB.

2.1.3 DMFE Switches

The DMFE has 7 mode-changing switches ($SW_{1\text{-}7}$) and 16 cascode DEM switches ($SW_{DEM1\text{-}16}$), as shown in Figure 2.7. Since the biasing current level of the front-end is 1nA, the non-linear pA-level leakage currents of these switches, and their spread, will introduce significant errors in both I_C and p.

In the pre-charge phase, SW_3, $SW_{4/6}$, and $SW_{5/7}$ are closed, while SW_1 and SW_2 are opened. The leakage currents due to the finite OFF resistance of SW_1 and SW_2 add additional currents to the biasing branches. In addition, the ON resistance of the chopper switches SW_4, SW_5, SW_6, and SW_7 affect the settling time of the bias amplifier. In the conversion phase, SW_3, SW_4, SW_6, SW_5, and SW_7 are opened, while SW_1, SW_2 are closed. The voltage drops caused by the ON-resistance of SW_1 and SW_2 adds to ΔV_{BE}. Whereas the leakage currents of switches SW_4, SW_5, SW_6, and SW_7 cause errors in p, leading to errors in ΔV_{BE}.

Moreover, in both pre-charge and conversion phases, the leakage currents of $SW_{DEM1\text{-}16}$ introduce errors in I_{bias}, because 7 ($=p$) switches (OFF) are connected to the branch biased at I_{bias}, while one switch (OFF) is connected to the branch biased at pI_{bias}, which causes errors in the current ratio, leading to errors in ΔV_{BE}.

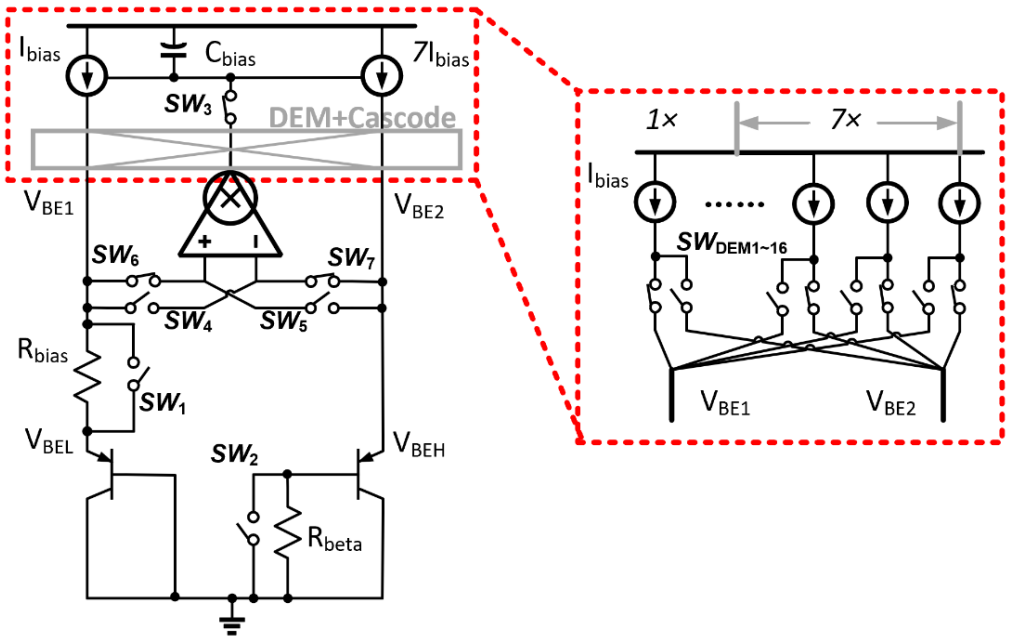


Figure 2.7 DMFE with dual-mode switches.

Calculating with Equations 2.1, 2.2, 2.3, and 2.6, the variations in leakage current (ΔI_{leak}), leakage current (I_{leak}), ON- and OFF-impedance (R_{ON} and R_{OFF}) requirements for each switch are shown in Table 2.1. The error contribution of each switch is limited to 0.03℃. For SW_1 and $SW_{4\text{-}7}$, since their input voltages are around V_{BE}, the R_{OFF} and ΔI_{leak} need to be designed carefully with similar requirements. For SW_2, since it is connected to the ground with a small voltage difference over it, its error can be easily limited to 0.01℃. For $SW_{DEM1\text{-}16}$, the requirement in ΔI_{leak} is the summation of 16 switches, especially for 8 OFF switches, therefore for each OFF switch, the limitation is 55fA. For SW_3, since I_{bias} is dependent on V_{bias} held on C_{bias}, its error, $\Delta I_{bias,error}$ is increasing with conversion time, leading to the error in V_{BE}. Thus, its requirement is determined by the conversion time (T_{conv}), the bias amplifier's transconductance ($g_{m,amplifier}$), and C_{bias}.

Table 2.1 Switch specifications in the DMFE.

Error types	Switches names	Error budgets	Error contributions
ΔV_{BE}	SW_1	$R_{ON} < 15k\Omega$ $R_{OFF} > 7G\Omega$ $\Delta I_{leak} < 450fA$	0.03℃
	$SW_{4\text{-}7}$	$R_{OFF} > 7G\Omega$ $\Delta I_{leak} < 450fA$	0.03℃
	SW_2	$R_{OFF} > 1G\Omega$ $R_{ON} < 8k\Omega$	0.01℃
	$SW_{DEM1\text{-}16}$	$\Delta I_{leak} < 450fA$	0.03℃
V_{BE}	SW_3	$I_{leak} < \frac{\Delta I_{bias,error}}{g_{m,amplifier}} \cdot \frac{C_{bias}}{T_{conv}}$ $\Delta I_{bias,error} < 2pA$	0.03℃

2.2 Tracking DSM readout

To realize BJT-based temperature sensors with high energy efficiency, a low-power tracking DSM has been proposed to digitize $X = V_{BE}/\Delta V_{BE}$ [13]. Compared to a zoom ADC [9], this requires a simpler digital controller, because it has no over-ranging requirements [14], which is suitable for a low-power sensor. In this design, the tracking DSM consists of a switched-capacitor DAC (SC-DAC), a discrete-time integrator, a 1-bit quantizer, and a Digital-Loop-Filter (DLF). As shown in Figure 2.8, V_{BE} is subtracted from the output of a multi-bit feedback DAC, $K\Delta V_{BE}$, which is accumulated with a gain of 'n'. The 1st SC integrator is followed by a comparator and a DLF. The DLF is the second integrator in digital domain with a feedforward compensation [15], and this will integrate 1-b bitstream to the multi-bit output with 1-bit step, which will achieve "tracking" by driving a feedback DAC [13]. D_{OUT} and K represent the same value, while K is the average of K_1 and K_2 as will be explained later.

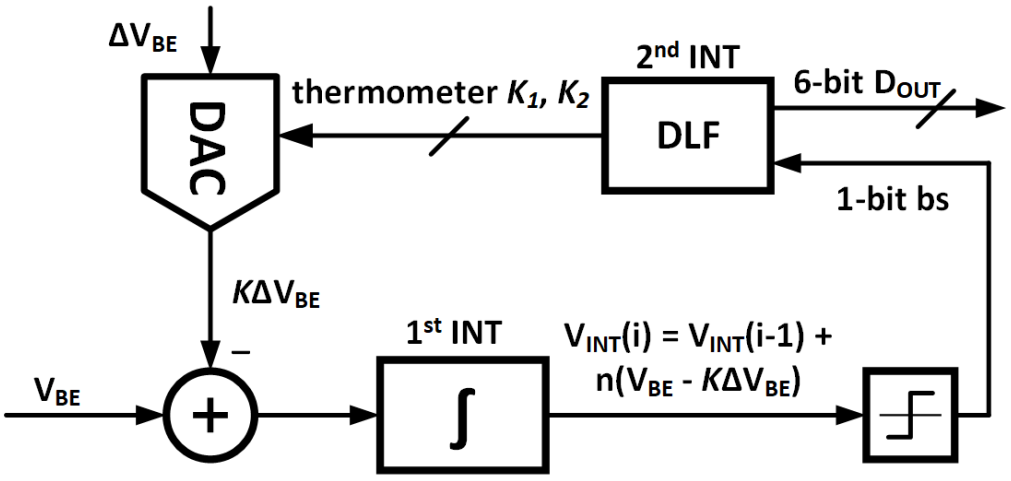


Figure 2.8 Tracking DSM topology.

To determine the range of D_{OUT}, its decimation result, $X = V_{BE}/\Delta V_{BE}$, is plotted in Figure 2.10 from -40℃ to 85℃ across different process corners. X varies from 18 to 6 when $p = 7$ and $I_{bias} = 1nA$. So, the range of D_{OUT} is chosen from 5 to 19 to realize stable DSM operation in the entire temperature range, corresponding to two 20-level thermometer codes, K_1, K_2, driving two halves of the SC-DAC, respectively, and a 6-bit D_{OUT}.

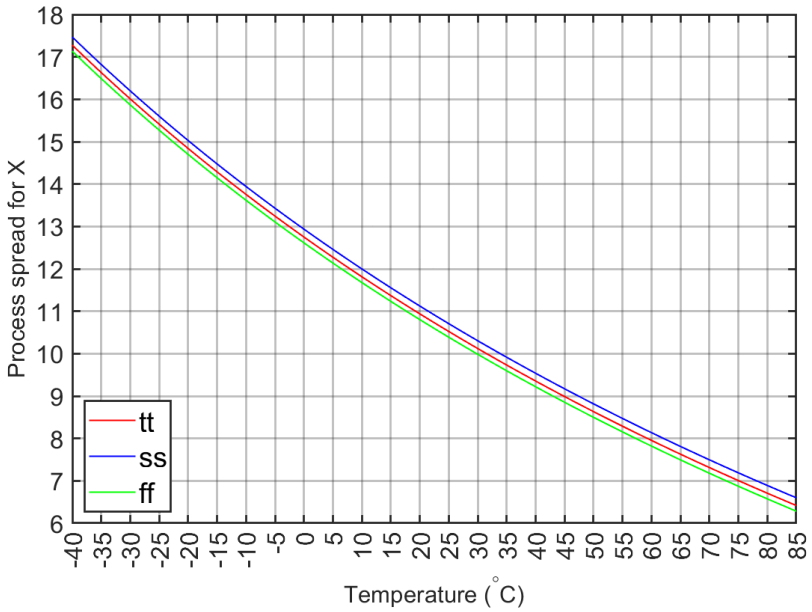


Figure 2.9 Decimated digital outputs *X* from -40℃ to 85℃.

The timing diagram of one conversion is shown in Figure 2.10, which consists of four phases: the bias phase (Φ_A), the initial phase (Φ_B), the tracking phase (Φ_C), and the DSM conversion phase (Φ_D). In Φ_A, the DMFE works in the pre-charge phase to generate I_{bias} while the readout circuit resets, resulting in a zero-output state. Then, the DMFE operates in the conversion phase. To relax the driving requirements of the front-end, Φ_B is necessary to pre-charge the SC-DAC elements to the mid code 12 before the remaining phases, while the rest of the readout circuit holds this initial state. The DSM starts to track the integrating voltage (V_{INT}) in Φ_C, by switching the DAC elements and the value of *K* is changed from the initial setting of 12 depending on the value of $V_{BE}/\Delta V_{BE}$. Finally, the DSM works in Φ_D, where the value of *K* toggles by LSB around an average value. In this phase, only one DAC element is switched (out of 20) in every DSM cycle. This significantly reduces the current required from the front-end to charge and discharge the 40 DAC elements.

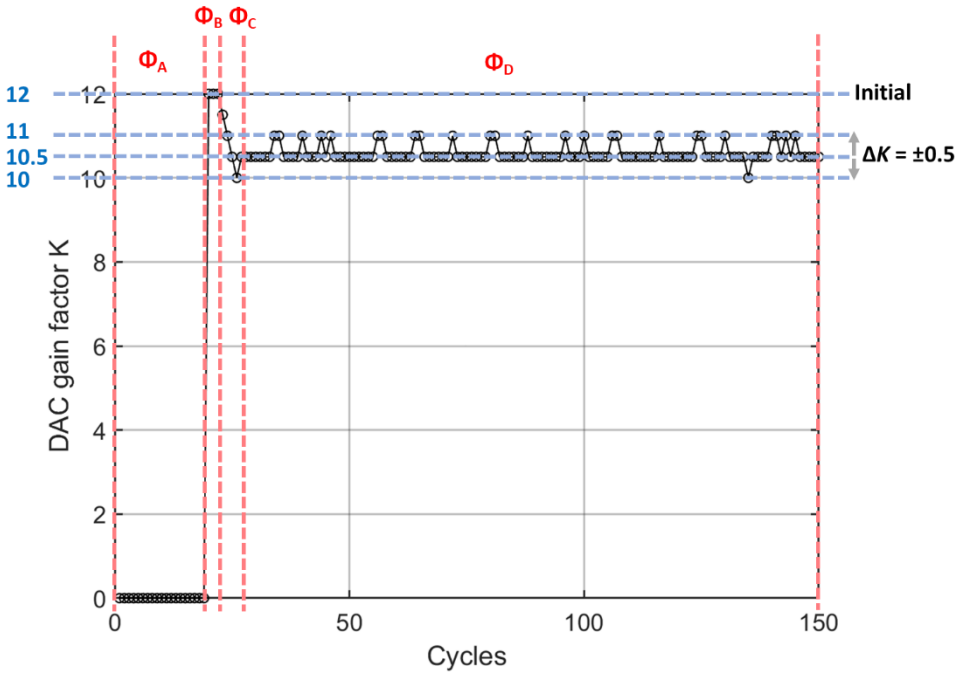


Figure 2.10 Timing diagram of tracking DSM and swing of DAC gain *K*.

The operation of the SC-DAC and integrator is shown in Figure 2.11, where C_S is the unit sampling capacitor, and C_{INT} is the integrating capacitor. In this design, both V_{BE} and ΔV_{BE} are simultaneously sampled and then integrated within one clock cycle. Two different gain factors, K_1 and K_2, can be applied to the two halves of the SC-DAC, making it possible to achieve 0.5-bit steps in K. For example, an average gain of $N = K + 0.5$, where N is an integer, can be obtained setting K_1 = N + 1 and K_2 = N, or by setting K_1 = N and K_2 = N + 1. This increases the DAC resolution by 2× without doubling the number of DAC capacitors [13]. However, this also introduces a $\pm 0.5\Delta V_{BE}$ common-mode (CM) variation. To avoid common-mode offset accumulation, cross-coupling feedback capacitors are used as the common-mode feedback (CMFB) [12].

Each main clock cycle is divided into two phases: the sampling phase (Φ_1) and the integration phase (Φ_2). In Φ_1, the integrator is auto-zeroed. V_{BEL} is sampled on a sampling capacitor, while $-\Delta V_{BE}$ is simultaneously sampled on K_1 and K_2 sampling capacitors. Thus, a charge proportional to $(V_{BEL} - 0.5(K_1 + K_2)\cdot\Delta V_{BE})$ is stored on the sampling capacitors. In Φ_2, the input voltages of the SC-DAC are swapped, and therefore, a charge proportional to $2\cdot(V_{BEL} - 0.5(K_1 + K_2)\cdot\Delta V_{BE})$ is integrated during each clock cycle.

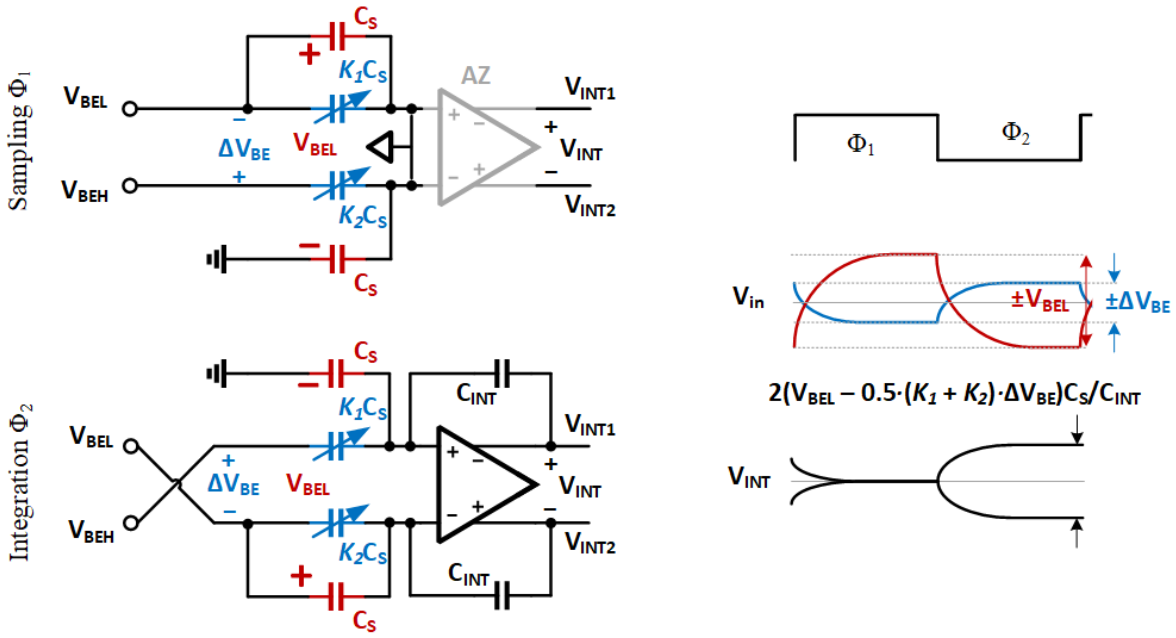


Figure 2.11 Operation of the SC-DAC and integrator in one clock cycle.

2.2.1 Oversampling ratio and thermal noise

For a targeted inaccuracy budget of ±0.15°C, the resolution should be 5 times better, i.e., 30mK (rms) at room temperature (20°C). For thermal noise-limited resolution, the quantization noise should then be below 7.5mK (rms), where SQNR = 84dB, which would require an OSR of 10000 with a conventional 1st order DSM, with a 1.2V reference. However, the output of the DAC of the tracking DSM switches between references of $K\Delta V_{BE}$ and $(K\pm0.5)\Delta V_{BE}$. As shown in Figure 2.12, an SQNR of 84dB can now be achieved with an OSR of 100, leading to a shorter conversion time. To achieve a thermal-noise level of 30mK (rms), a sampling capacitor of 820fF ($C_{S,total}$ = 41fF × 20) is required for an OSR of 100, which is 12dB lower than the SQNR.

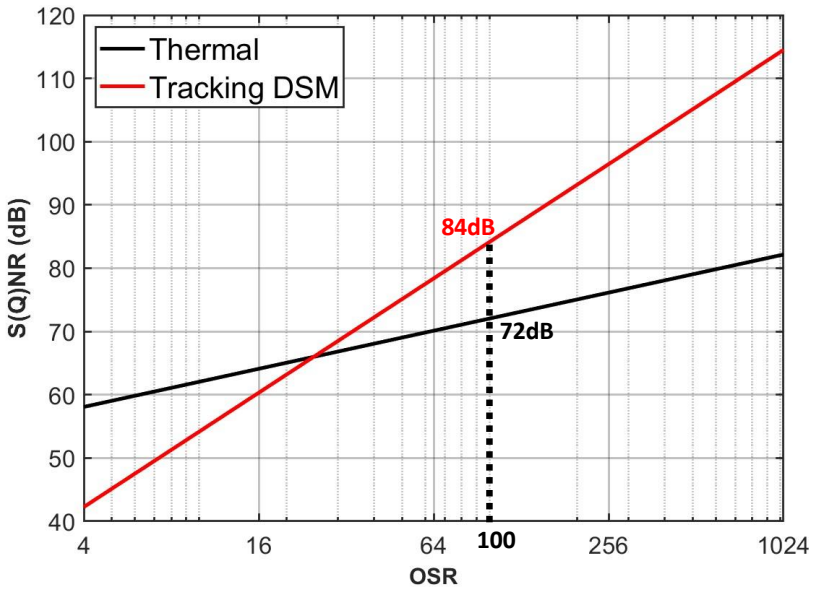


Figure 2.12 Comparison between SNR of kT/C noise and SQNR of the DSM over OSR

2.2.2 Driving capability of the front-end

To achieve the operation described in Figure 2.11, the front-end alternately generates V_{BE1} and V_{BE2} during Φ_1 and Φ_2. Since I_{bias} is only 700pA at -40 °C, settling errors in the sampled voltages V_{BE} and ΔV_{BE} may occur, as shown in Figure 2.13. For an error budget of 0.03 °C, the error caused by ΔV_{BE} needs to be less than 4.8μV. Since the settling errors in ΔV_{BE} have 20% spread, such errors should be less than 4.8μV/(20%·0.5ΔV_{BE}) ~ 0.01%, which corresponds to a ~7τ settling. Since the output resistance of the front-end ($R_{OUT,FE}$) is about 37MΩ (≈V_T/I_C) [4], 7τ is 0.2ms, where $\tau = R_{OUT,FE} \cdot C_{S,total}$. Since the sampling period is half of a clock cycle, the maximum sampling frequency is 2.5kHz, corresponding to the minimum conversion time of 80ms.

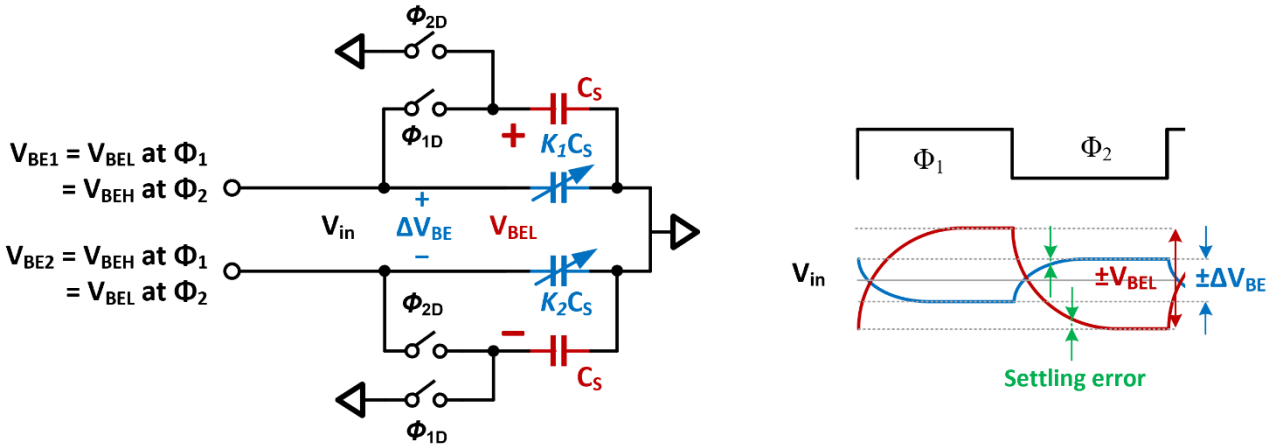


Figure 2.13 Settling behavior of the sampling capacitors.

2.2.3 Sampling switches leakage considerations

As shown in Figure 2.14, since 40 sampling switches are implemented to switch the 40 unit DAC elements (20 in each differential half), the leakage currents in the conversion phase is the largest error source compared to all the leakage discussed in the DMFE. A larger error budget of ±0.05°C for the readout sampling switches is needed. Thus, the variation in p should be less than 0.038% as calculated from Equation 2.4, which corresponds to a maximum allowed variation of ±2.66pA in the leakage current (±66fA for each sampling switch).

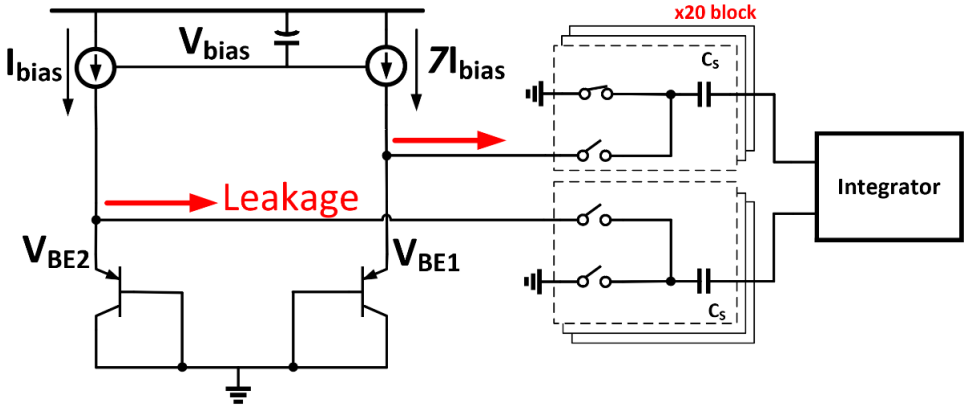


Figure 2.14 Sampling switches connected to the DMFE.

2.3 Error correction techniques

Besides the temperature errors discussed above, the mismatch of PNPs, current mirrors, and sampling capacitors will also cause error. Various dynamic techniques will be used to minimize the effect of mismatch, as well as to reduce 1/*f* noise. In addition, a PTAT trimming technique will be used to compensate for the PTAT spread of V_{BE} in the digital domain.

2.3.1 DEM for BJT transistors, current mirrors, and sampling capacitors

In this design, to get good accuracy, the errors sensed in the conversion phase must be reduced. In the pre-charge phase, the spread of the PNPs causes I_{bias} to spread, which, in turn, creates errors in V_{BE} during the conversion phase. Without trimming, the mismatch and process spread of the PNPs will cause ±0.2°C inaccuracy. With a PTAT trim, as will be mentioned later, this error can be further reduced to ±0.02°C. Thus, a DEM of PNPs during the pre-charge phase is not necessary.

However, in the conversion phase, the mismatch of the PNPs will cause causes current density errors that directly affect ΔV_{BE}, and thus introduce significant errors. Therefore, the PNPs are swapped to mitigate such errors, as shown in Figure 2.15. This also ensures that V_{BE} switches between VBEH and VBEL during Φ_1 and Φ_2, as described in Figure 2.13. In Φ_1, Q1 is biased with I_{bias} while Q2 is biased with $7I_{bias}$. In Φ_2, the biasing currents flowing through Q1 and Q2 are swapped to $7I_{bias}$ and I_{bias}, respectively. This modulates the mismatch errors in ΔV_{BE} to the DEM frequency, which is equal to the sampling frequency (f_S).

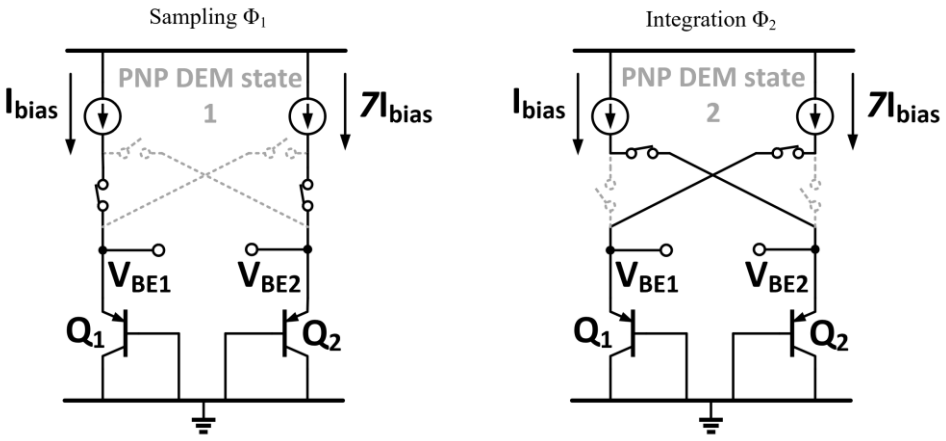
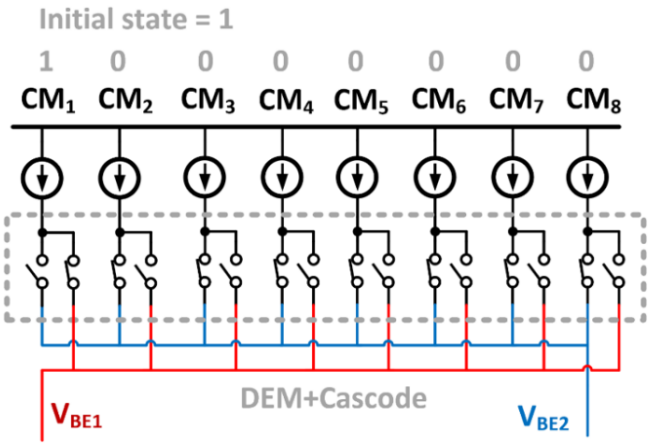


Figure 2.15 DEM of PNPs in the conversion phase.

Similarly, in the pre-charge phase, the spread of the current mirrors results in an error in I_{bias}, affecting V_{BE} in the conversion phase. To mitigate this error, the current mirrors are implemented with large transistors. In the conversion phase, the spread of 8 current mirrors $CM_{1\text{-}8}$ will cause errors in ΔV_{BE}. To mitigate this, a barrel shifting DEM has been implemented as shown in Figure 2.16. When the state of CM is 1, the current mirror is connected to V_{BE1}, while the state is 0, the current mirror is connected to V_{BE2}. This is synchronized with the DEM of PNP for Φ_1 and Φ_2. The DEM frequency is chosen to be $f_S/8$ (one complete cycle).



CM DEM state	$CM_{1\text{-}8}$ (Φ_1)	$CM_{1\text{-}8}$ (Φ_2)
1 (initial)	10000000	01111111
2	01000000	10111111
3	00100000	11011111
4	00010000	11101111
5	00001000	11110111
6	00000100	11111011
7	00000010	11111101
8	00000001	11111110

Figure 2.16 DEM of current mirrors.

The mismatch between the SC-DAC elements causes an error in K (PTAT voltage gain), which varies over temperature. This affects the decimated value X, which further introduces temperature errors after transferring to μ. As shown in Figure 2.17, a barrel-shifting DEM among 20 capacitors for each half of the SC-DAC has been applied to mitigate the mismatch-introduced gain errors.

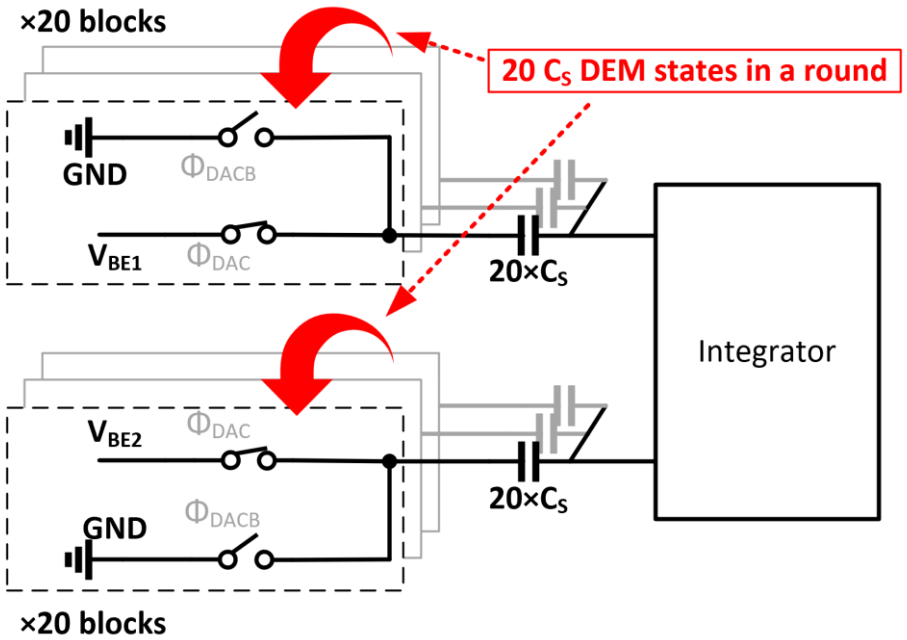


Figure 2.17 DEM of sampling capacitors.

When barrel-shifting DEMs are applied to the SC-DAC and current mirrors, the intermodulation between the output bitstream and DEMed-mismatch signal around $f_S/2$ will be folded back to signal band, as in [16]. This undesired folded signal will probably affect the resolution by raising the noise floor. To avoid this effect, the mismatch of the SC-DAC and current mirrors need to be well controlled.

2.3.2 System-level chopping

Even after the use of DEM, mismatch-introduced errors in the bias amplifier, and the quantizer still exist and will limit accuracy. To mitigate this, system-level chopping (Φ_{CHL}) is incorporated as shown in Figure 2.18. After every DSM conversion period (100 clock cycles), the input chopper switches the polarity of (V_{BE1}-V_{BE2}), while the output chopper is implemented in the digital domain and switches the polarity of the output bitstream. Thus, two different decimated outputs, *X*, are generated, and the residual offset errors are mitigated by averaging these outputs. Since the system-level chopping only mitigates errors in the conversion phase, the offset error in the bias amplifier in the pre-charge phase still exists. So, this amplifier is chopped after every DSM conversion period by resetting V_{bias} in the pre-charge phase, at the same frequency as Φ_{CHL}.

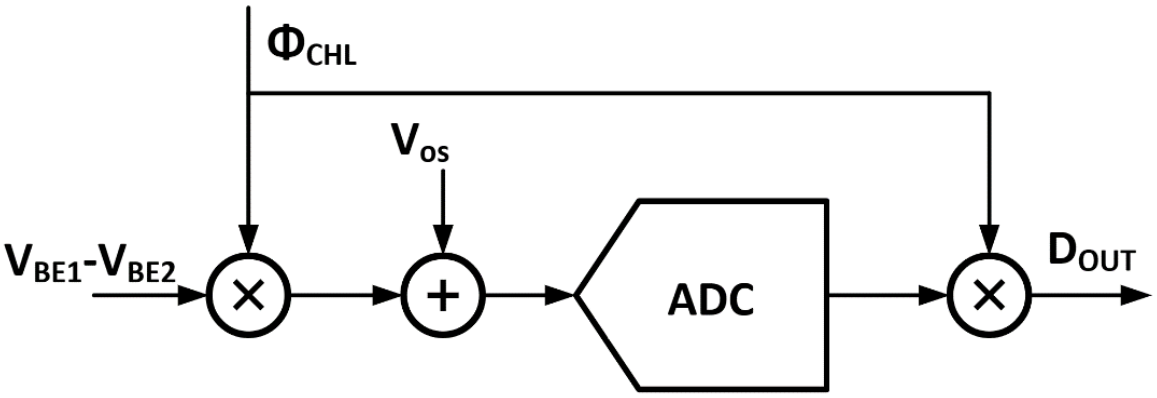


Figure 2.18 System-level slow chopping.

The system with all the dynamic techniques is shown in Figure 2.19. The timing diagram of DEM and system-level slow chopping are shown in Figure 2.20. In the pre-charge phase, the Φ_{CHL} signal only controls the bias amplifier by Chopper1, while CM and PNP DEM techniques are kept in the reset state. In the conversion phase, CM and PNP DEM techniques are operated with Φ_{CHL}, denoted as Chopper2. When the Φ_{CHL} is high, PNP DEM starts with state 1, representing that V_{BE1} equals V_{BEL}. When the Φ_{CHL} is low, PNP DEM starts with state 2, representing that V_{BE1} equals V_{BEH}. Additionally, the polarity of the comparator is changed with the Φ_{CHL} by Chopper3. These techniques and specifications limit the trimmed error to ±0.15°C from -40°C to 85°C.

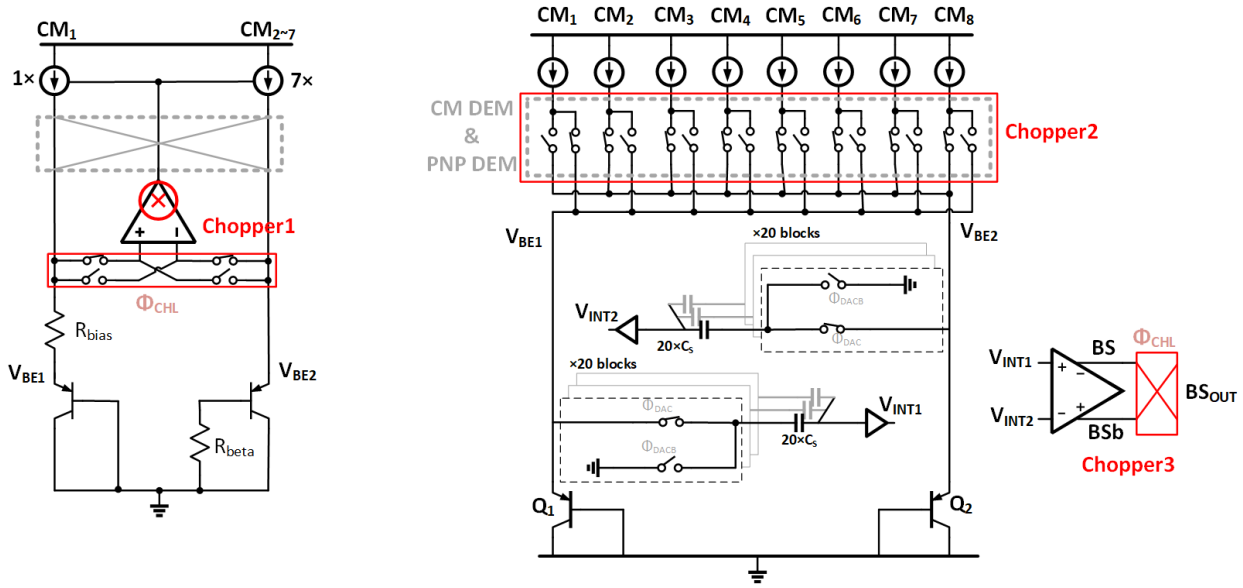


Figure 2.19 System topology with dynamic techniques: pre-charge phase (left), conversion phase (right).

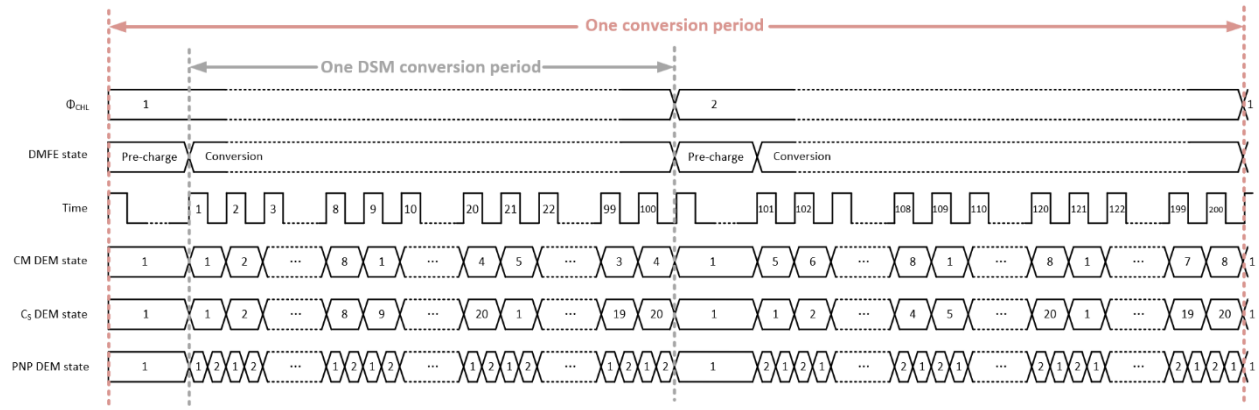


Figure 2.20 Time diagram of DEM and Φ_{CHL}.

2.3.3 Trimming techniques

To get the temperature information from this sensor, the decimated output X needs to transfer to a temperature-linear output μ, which can be expressed as [17]:

$$\mu = \frac{\alpha \cdot \Delta V_{BE}}{\alpha \cdot \Delta V_{BE} + V_{BE}} = \frac{\alpha}{\alpha + V_{BE}/\Delta V_{BE}} = \frac{\alpha}{\alpha + X}, \tag{2.16}$$

As mentioned earlier, the process spread in V_{BE} requires a room temperature PTAT digital trim. Since the offset spread of decimated output X results in the PTAT spread in temperature [17], as shown in Figure 2.21, which means applying an offset trim to X can correct the PTAT errors.

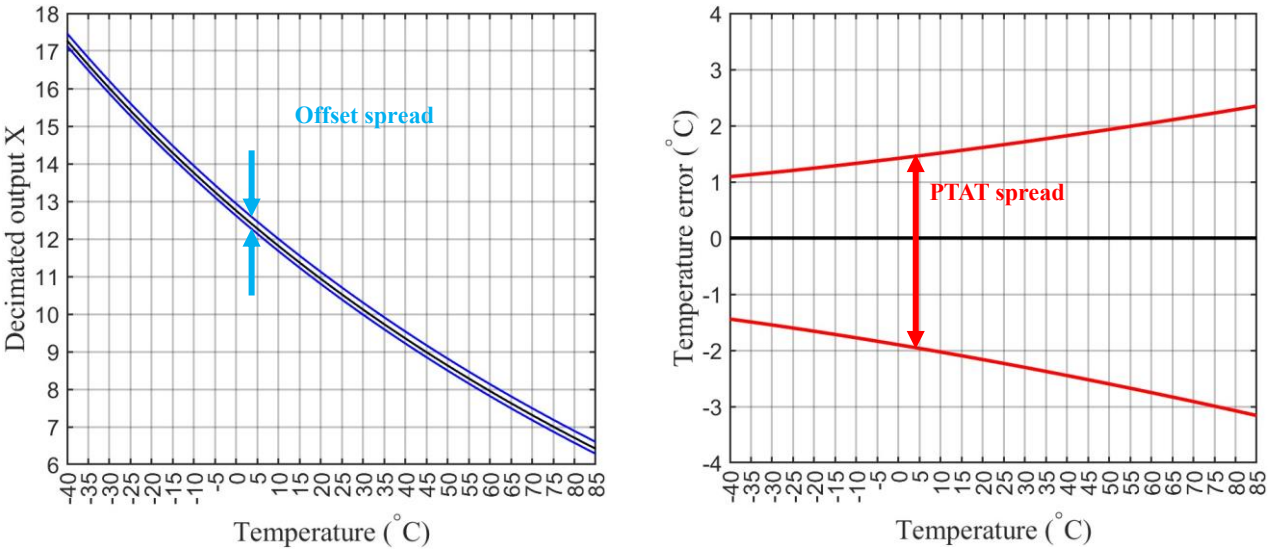


Figure 2.21 A offset spread (left) of the decimated output X, a PTAT spread (right) of temperature (α = 15)

2.4 Summary

In this chapter, the architectural design of a front-end and readout has been presented which can meet the required targets for accuracy (<±0.15°C) and resolution (<30mK) with the minimum power. Different error sources due to leakage, process spread and component mismatch have been analyzed and various techniques have been applied to mitigate them. Error budgets weree made to meet the target accuracy and lay the foundations for the circuit-level implementations.

3 Circuit implementation

This chapter discusses the transistor-level circuit implementation of the proposed PNP-based sensor, based on the design considerations discussed in the previous chapter. Two main circuit blocks are discussed in detail: the DMFE and the tracking DSM. Finally, the power breakdown is presented.

3.1 BJT Dual-Mode Front-End

The PNP-based DMFE can be divided into four main sub-circuits, which are highlighted in Figure 3.1: the current mirrors, the bias amplifier, the BJTs, the resistors, and the sampling switches. Φ_{AFE}, Φ_{AFED}, and Φ_{AFEBD} are the signals used to control the circuit's operating phases, as mentioned in Figure 2.1. The following sections discuss these sub-circuits separately.

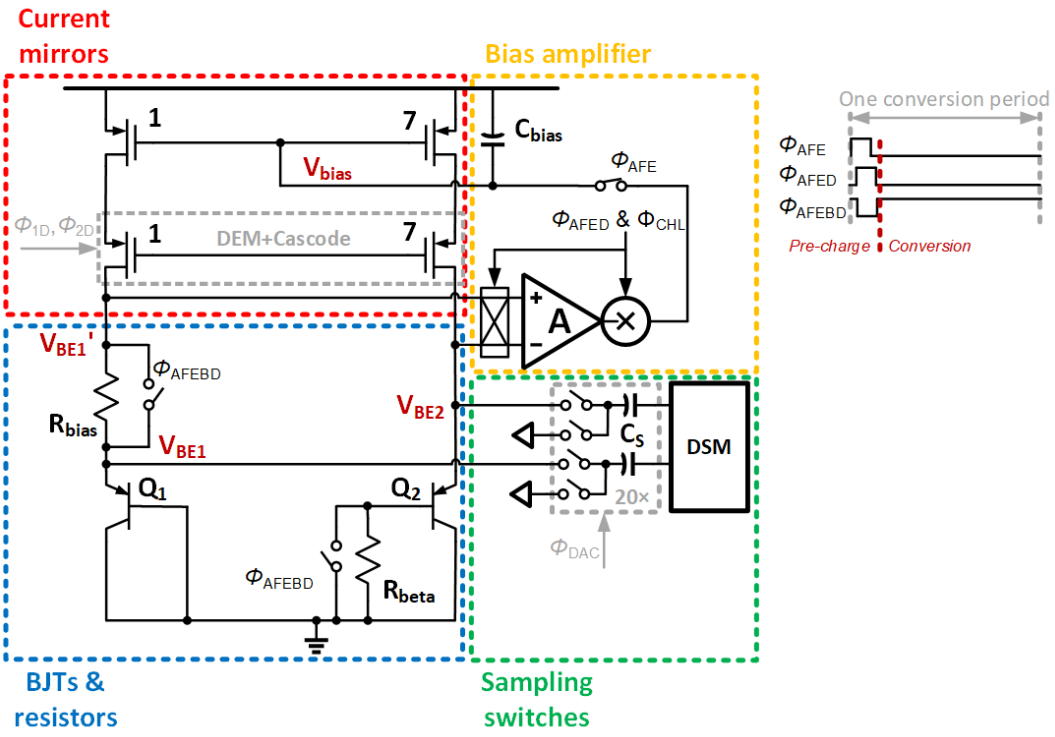


Figure 3.1 Sub-circuits of the PNP-based DMFE (left), and its timing diagram (right).

3.1.1 BJTs and resistors

The BJTs and resistors sub-circuit highlighted in blue in Figure 3.1, consists of R_{bias}, R_{beta}, some switches, and the PNPs. The key design considerations are the leakage currents in switches and the non-idealities in R_{bias} and R_{beta}, which introduce errors in V_{BE} and ΔV_{BE}. As discussed in Chapter 2.2.1, 2μm×2μm substrate PNPs minimize I_S non-linearity at low bias currents. To achieve I_{bias} = 1nA, when p = 7, R_{bias} and R_{beta} are set to 56MΩ and 8MΩ respectively.

To get good accuracy, the spread of the resistors needs to be considered. In the targeted TSMC 180nm process, the high-resistance p-poly resistor (hipo resistor), which has a high sheet resistance, is the best choice for low area. Simulations show that its TC spread causes ±0.02°C inaccuracy.

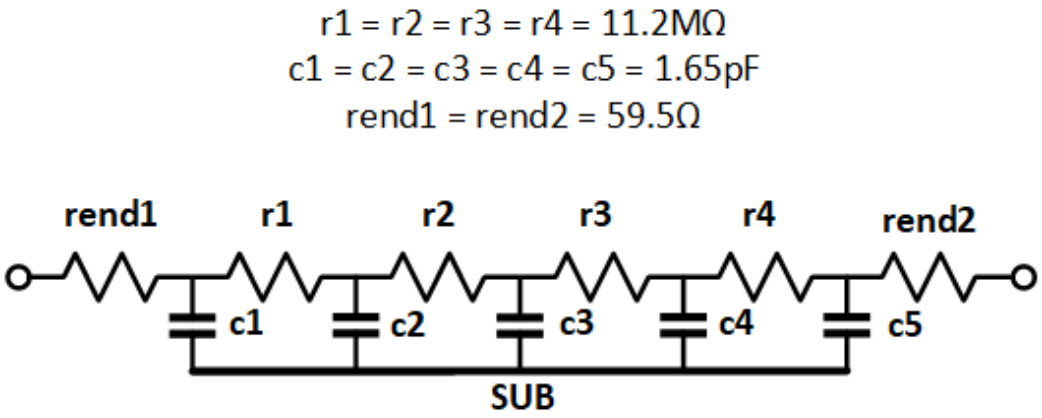


Figure 3.2 A 56MΩ (W/L = 1μm/1mm) p-poly resistor model.

As shown in Figure 3.2, R_{bias} has a large distributed parasitic capacitance (C_{par} ~ 8pF) to the substrate. This causes an additional pole in the feedback loop of the bias amplifier, as shown in Figure 3.3, which, in turn, causes instability during the pre-charge phase. There are two dominant poles: a first pole (ω_{p1}) determined by C_{bias}, and a second pole (ω_{p2}) determined by C_{par}. To stabilize the feedback loop, ω_{p1} is reduced by increasing C_{bias}, at the expense of slower settling. Since the second pole ω_{p2} caused by C_{par} is at ~250Hz, the unity gain bandwidth is reduced to 125Hz by increasing C_{bias} to 60pF, to ensure a phase margin >60°.

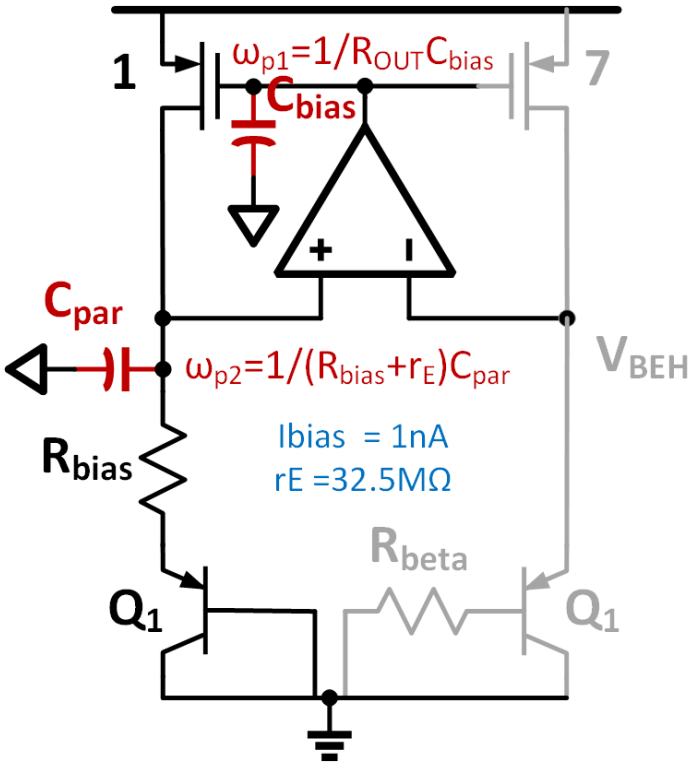


Figure 3.3 Two poles of the DMFE with circuit implementation.

As shown in Figure 3.4 (right), over the intended temperature range (-40°C to 85°C), V_{BE1}' varies from 740mV to 440mV, while ΔV_{BE} varies from 40mV to 60mV. SW_1 is implemented as a 0.22μm/0.18μm Regular-Threshold-Voltage (RVT) NMOS, which when driven by a boot-strapped driving voltage 0-2V_{DD}, achieves 8kΩ ON-resistance, negligible bulk leakage variations, and 20GΩ OFF-resistance. For SW_2, whose drain and source voltages are close to ground, a 0.22μm/0.35μm High-Threshold-Voltage (HVT) NMOS is used to realize an OFF-resistance of 8GΩ.

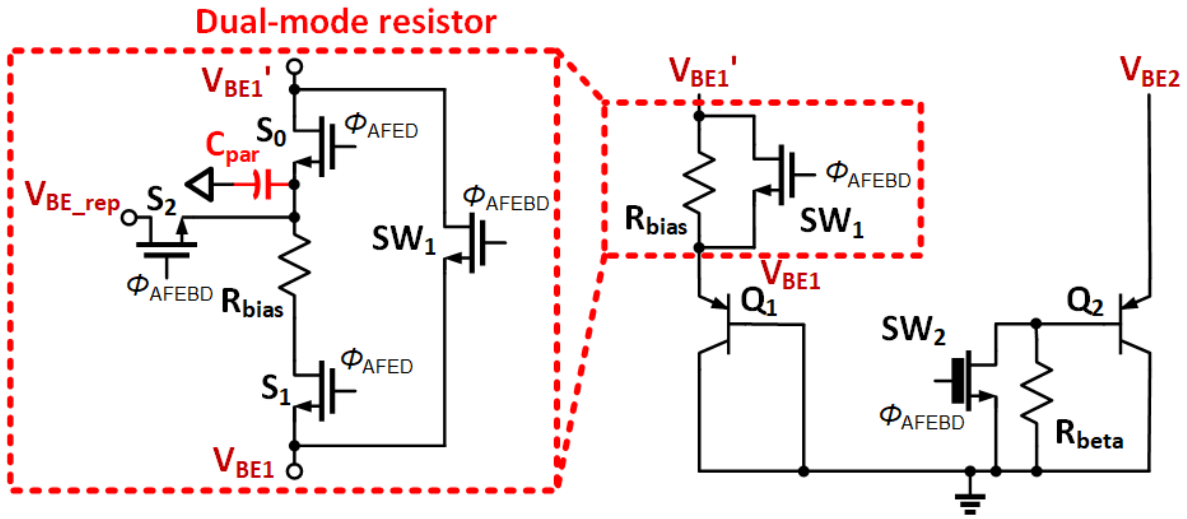


Figure 3.4 A dual-mode resistor (left), switches implementation in the resistors and BJTs block (right).

As mentioned in Figure 2.15, DEM is implemented by swapping the current flowing into the two PNPs for each clock cycle. As shown in Figure 3.5 (left), when I_{bias1} varies from $7I_{bias}$ to I_{bias}, due to C_{par}, V_{BE1}' settles much slower than V_{BE2}, resulting in a dynamic error in ΔV_{BE}. To prevent this, a dual-mode bias resistor is proposed to isolate C_{par}, as shown in Figure 3.5 (right). In the conversion phase, S_0 and S_1 are switched OFF, while S_2 is switched ON to keep the voltage on C_{par} by connecting to a replica voltage of V_{BEL} (V_{BE_rep}) generated by another replica branch. As shown in Figure 3.4 (left), S_0, S_1, and S_2 are all 0.22μm/0.18μm NMOSs with 0-2V_{DD} driving voltage to minimize the ON-resistance.

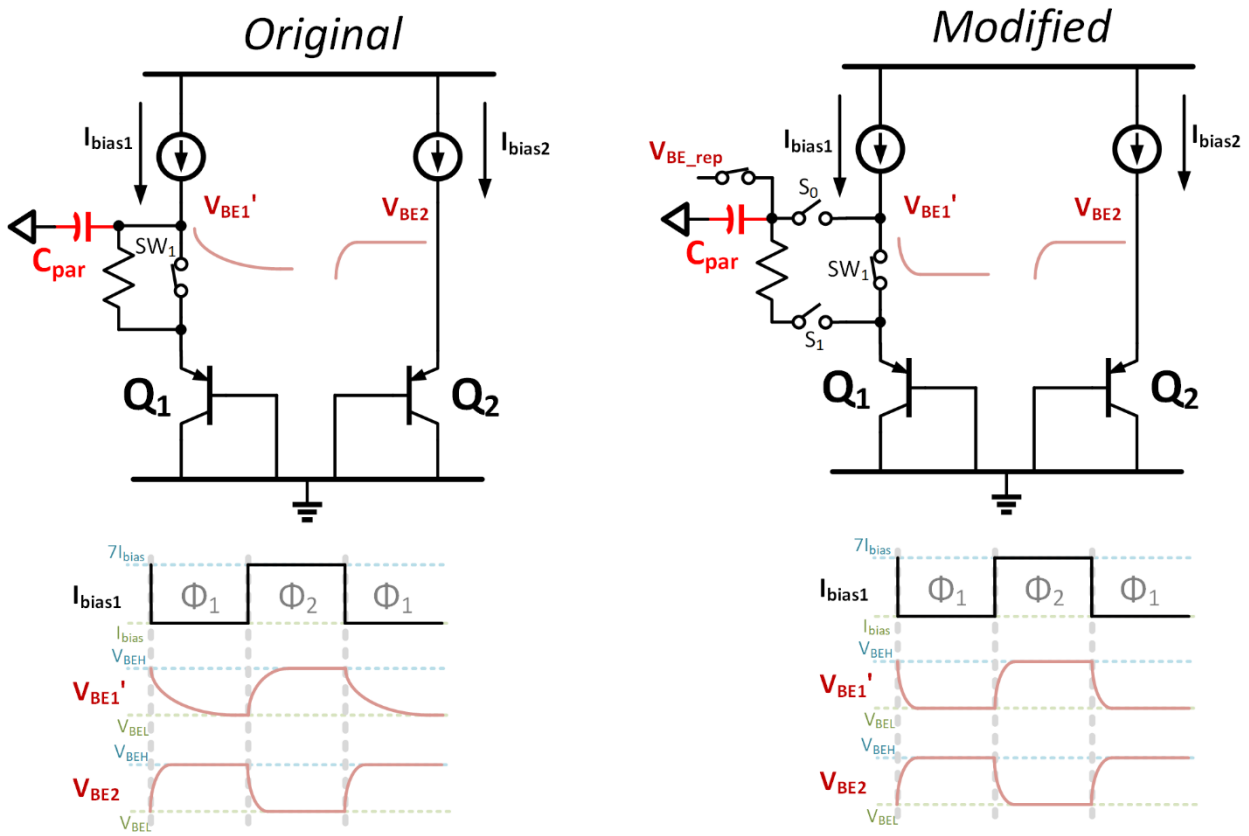


Figure 3.5 A Modified R_{bias} to mitigate the settling problem caused by C_{par}.

3.1.2 Bias generator

The current mirrors and cascode stage are highlighted in red sub-circuit of Figure 3.1, which will be used in the bias generator. Three important considerations should be kept in mind: the mismatch among the current mirrors in the pre-charge phase, the leakage current flowing through the cascode stage, and the gate capacitor.

3.1.2.1 Current mirrors and additional biasing branch

The circuit diagram of the current mirrors along with the additional biasing branches (to bias cascode devices and amplifier) is shown in Figure 3.6. Since DEM is not applied during the pre-charge phase, the size of the PMOS current mirrors ($M_{1\text{-}8}$) is mainly determined by the desired mismatch, which should result in less than ±0.03°C (3σ) error in the V_{BE}, used during the conversion phase. As a result, they are biased in strong inversion, with $g_m/I_d = 10$ and W=2μm, L=120μm. To improve the front-end's power supply rejection, $M_{1\text{-}8}$ are cascoded. To minimize their source/drain leakage currents, the cascode transistors are small, with W=0.5μm, L=0.5μm. And an additional biasing branch generates the biasing voltages V_{BN}, V_{CN}, and V_{CP} required by the bias amplifier and the integrator. A start-up circuit is also required to ensure the desired non-zero operating point. During start-up, V_{bias} is briefly pulled up to $V_{DD}/2$ by M_{S1}, which is driven by an external control signal.

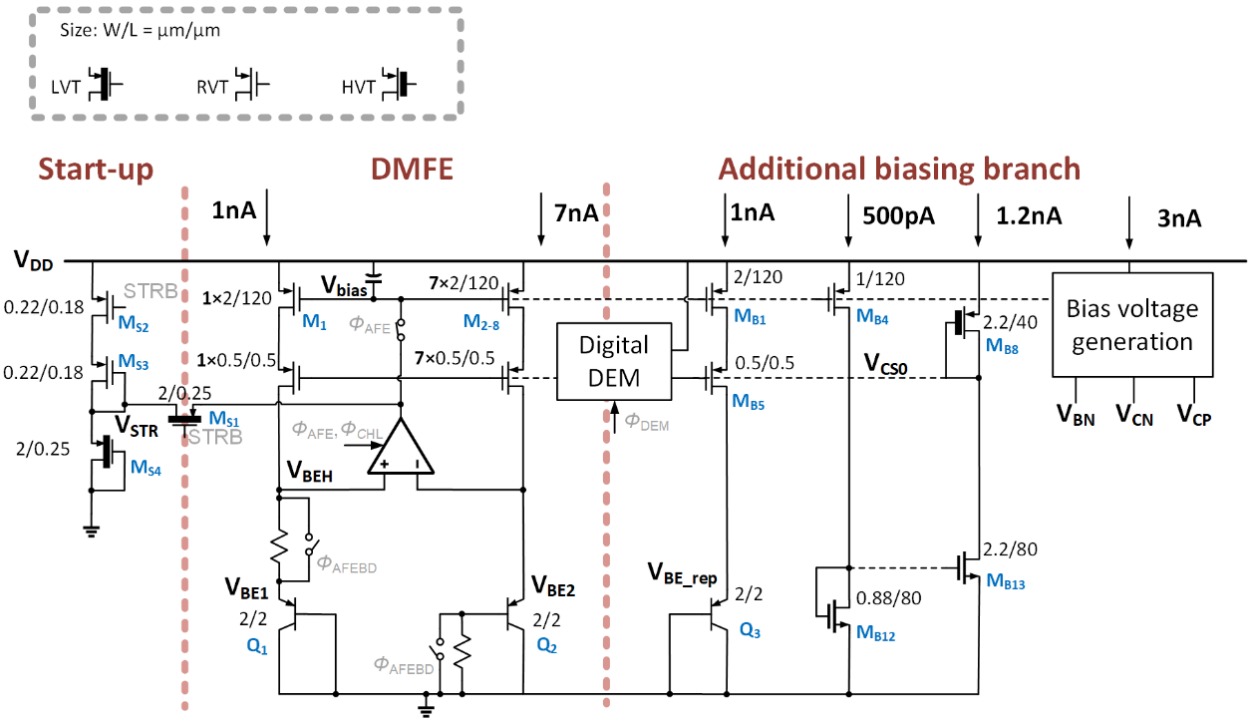


Figure 3.6 Schematic of the front-end.

3.1.2.2 DEM circuitry

As described in Chapter 2.4.1, the mismatch of the current mirrors will cause errors in the ΔV_{BE} generated during the conversion phase. To mitigate this, DEM has been applied to the current sources. As shown in Figure 3.7, the cascode transistors are re-used as DEM switches [13]. For example, when Φ_{DEM} is 1, M_9 is turned ON by connecting its gate to V_{CS0}, while M_{17} is turned OFF by connecting its gate to V_{DD}. This connects the unit current source M_1 to Q_1 to generate V_{BE1}. When Φ_{DEM} is 0, M_{17} is turned ON, while M_9 is turned OFF, connecting M_1 (together with 6 other current sources) to Q_2 to generate V_{BE2}.

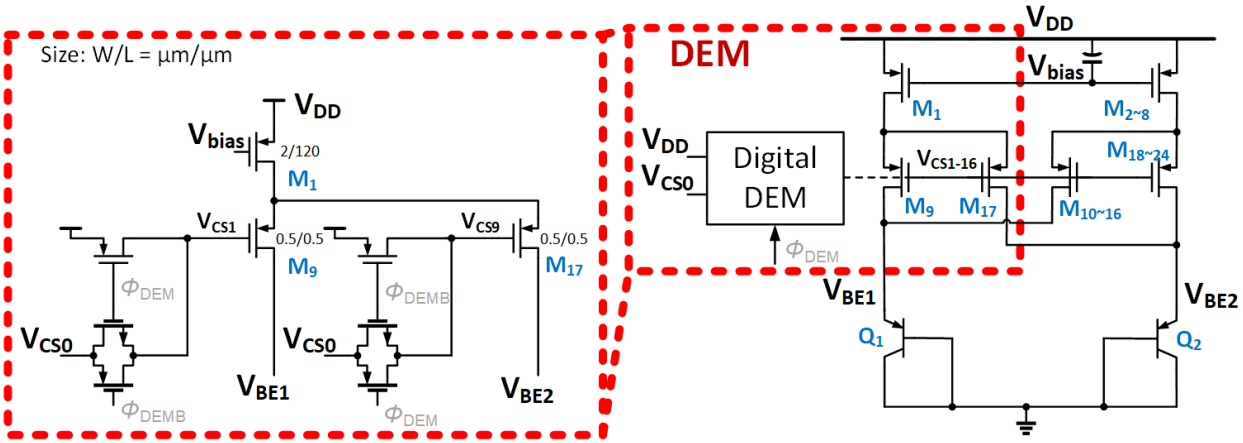


Figure 3.7 Digital DEM circuitry with switches in the conversion phase.

However, the DEM transitions cause spikes on V_{CS0}, which may cause settling errors during the conversion phase. To switch the currents going to Q_1 and Q_2, the gate voltages (V_{CS}) of the cascode transistors must be switched between V_{CS0} and V_{DD}. As shown in Figure 3.8, when V_{CS} is switched from V_{DD} to V_{CS0}, the gate capacitances of $M_{9\text{-}24}$ must be pulled down by the V_{CS0} generator (M_{B8} in Figure 3.6). This results in a spike (V_{spike}) on V_{CS0} after every DEM state change. During a Φ_1 to Φ_2 transition, this spike is caused by the $M_{17\text{-}24}$ transistors, whereas, during Φ_2 to Φ_1 transitions it is caused by the $M_{9\text{-}16}$ transistors. Due to the nA-level biasing currents, however, V_{CS0} may not settle completely back to $V_{CS,ideal}$ within a half clock cycle. The incomplete settling of V_{CS} then causes errors in the V_{BE} sampled by the readout.

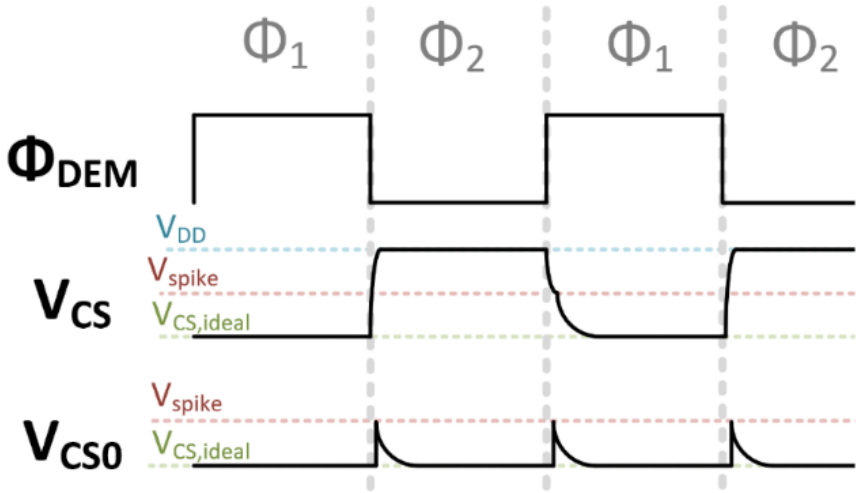


Figure 3.8 Timing diagram of cascode voltage settling.

Two measures are implemented to mitigate the V_{CS0} spikes. First, the settling time of the spikes is reduced by increasing the current in the V_{CS0} generating bias branch to 1.2nA. Second, the peak voltage V_{spike} of the spikes is reduced by connecting a decoupling capacitor to the gate of M_{B8}. These measures reduce the extra V_{BE} error to less than 0.02°C.

3.1.3 Bias amplifier and capacitor

In this section, the design considerations of the bias amplifier, switch, and C_{bias} (the yellow sub-circuit in Figure 3.1,) are discussed. The bias amplifier suffers from offset, noise, finite gain, and limited gain-bandwidth product, while the size of C_{bias} is determined by the leakage current flowing through SW_3.

As calculated in Chapter 2, the open-loop gain of the bias amplifier should be larger than 58dB. To achieve this, it is implemented as a folded-cascode amplifier, as shown in Figure 3.9. Its open-loop gain is greater than 71dB (over PVT), resulting in temperature errors less than 0.01℃. Since I_{bias} errors in the pre-charge phase cause V_{BE} errors during the conversion phase, chopping is used to mitigate the amplifier's offset. Since the input voltage of the bias amplifier is V_{BE} (ranging from 440mV to 740mV), and V_{DD} ~ 1V, the amplifier employs an NMOS input pair, while the input chopper switches are implemented with bulk-source-connected PMOS transistors for their lower ON-resistance (8kΩ). To reduce the bulk leakage at high temperatures, the chopping switches M_{CHP1-4} are small (0.18μm/0.22μm).

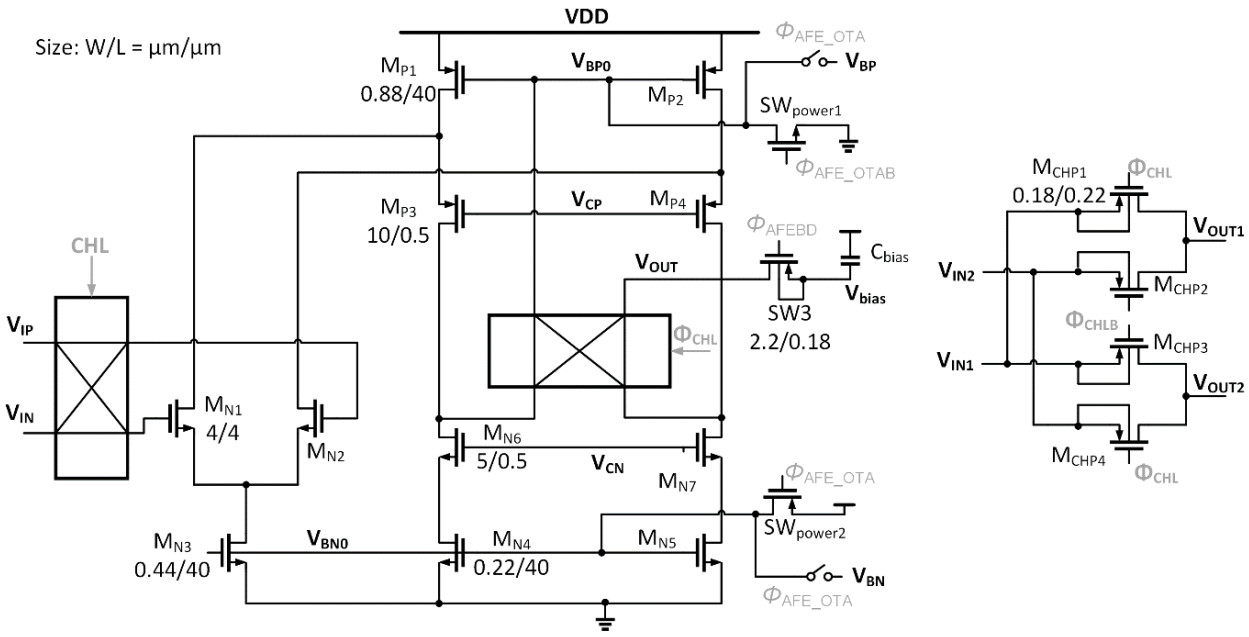


Figure 3.9 Schematic of the bias amplifier (left) and chopper (right)

In the pre-charge phase (Φ_{AFE_OTA} = 1), the bias amplifier draws 2.5nA in each branch, for a total of 10nA. M_{N1} and M_{N2} are sized 4μm/4μm to keep the input pair operating in weak inversion region, resulting in a transconductance ($g_{m,amplifier}$) of 50nS. As discussed in , for stability, the I_{bias} generating loop must have a unity gain frequency $g_{m,amplifier}/C_{bias}$ of less than 125Hz. This was achieved by setting C_{bias} to 60pF, resulting in a settling time of about 6ms. In the conversion phase, the bias amplifier is disabled by connecting V_{BP} to the ground and connecting V_{BN} to V_{DD} via SW_{power1} and SW_{power2} (Φ_{AFE_OTA} = 0). C_{bias} is a MIM capacitor, which can be placed on top of the large p-poly Rbias to save area. SW_3 is implemented as a bulk-source-connected PMOS (2.2μm/0.18μm), to obtain a small variation in leakage current and a low ON-resistance. This choice introduces less than 0.01℃ temperature error.

3.1.4 Sampling Switches

There are in total 20 unit DAC elements in each half of the differential SC-DAC as highlighted in the green sub-circuit in Figure 3.1. As shown in Figure 3.5, each unit DAC element has a switch (SW_{sama}) connected to the front-end and another switch (SW_{samb}) connected to the ground. The control signals $\Phi_{DAC1-40}$ consist of thermometer codes K_1 and K_2. Their leakage currents and the influence of their ON-resistance on settling speed are the key considerations for designing these switches.

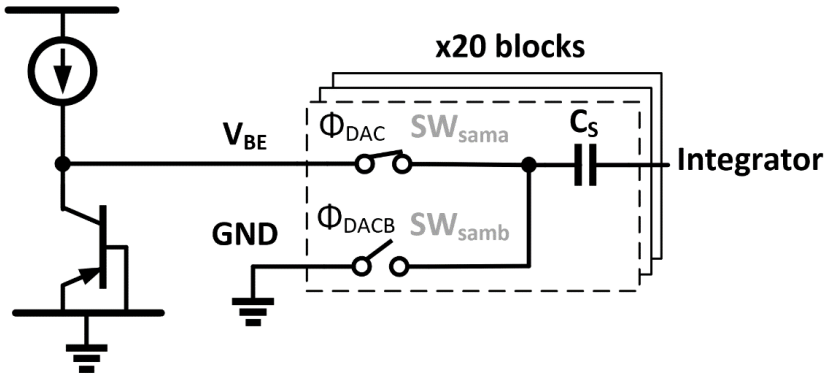


Figure 3.10 Sampling switches block topology for half SC-DAC.

To reduce the error caused by the leakage currents, the variation of the drain-source and bulk leakage currents over temperature need to be reduced. From -40°C to 85°C, V_{BE} varies from 740mV to 440mV. The resulting variation in the drain-source leakage currents, is more significant than that in the bulk leakage currents. In [13], T-switches were used to minimize the leakage current flowing out of the front-end. Their center tap was connected to a replica bias voltage V_{BEH_R}, as shown in Figure 3.11 (left). When the SW_{sama} are OFF (Φ_{DAC} = 0), its mid points are connected to V_{BEH_R}. However, this results in a large variation in the drain-source leakage of M0, leading to ±0.1°C error. This is mainly due to the variation of V_{BEH_R}, which, in turn, is caused by the variation in R_{bias} and in the leakage current entering the T-switch ($I_{leak,VBER}$). In this work, to achieve less than ±0.05°C error, a modified unit DAC element is proposed, as shown in Figure 3.11 (right). M0 now has a bulk-source connection, and the replica bias voltage is V_{BEL}, which lowers the resulting leakage current and minimizes its variation, as will be explained below.

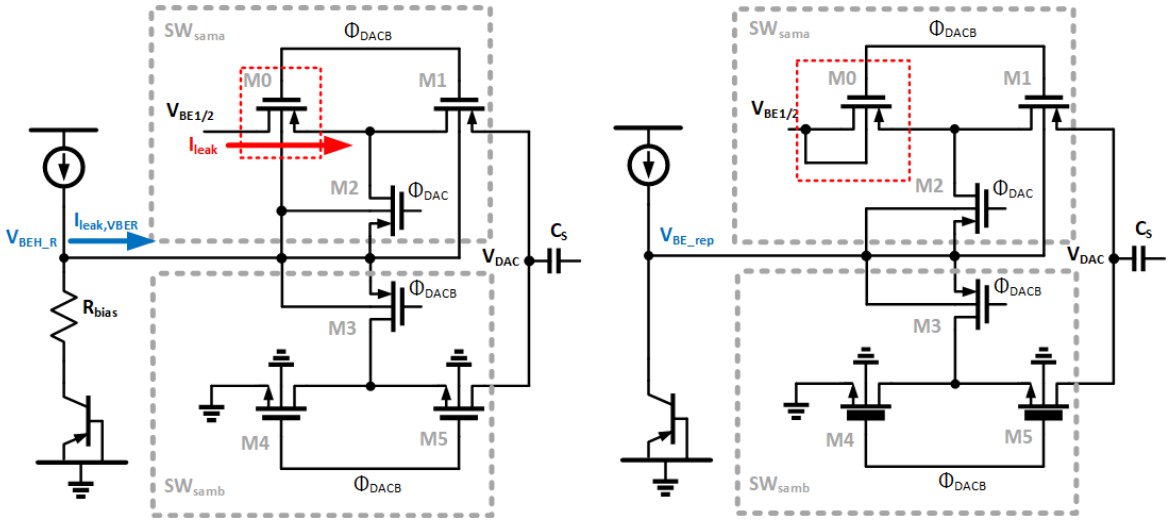


Figure 3.11 Schematic of a unit DAC element with T-switches in [13] (left), and a modified one (right).

Two phases of a unit DAC element are shown in Figure 3.12. When Φ_{DAC} is high, SW_{sama} is ON to connect V_{DAC} of C_S with V_{BE}, while SW_{samb} is OFF. In this case, M2 is turned OFF, and V_{DAC} is charging to V_{BE} through M0 and M1. The leakage current flowing out from $V_{BE1/2}$ mostly depends on the drain-source leakage of M4 and M5. To reduce that, M4 and M5 are both HVT NMOSs, while the source voltage of M5 is charged to V_{BE_rep}. When Φ_{DAC} is low, C_S is connected to the ground by switching ON SW_{samb} and switching OFF SW_{sama}. As mentioned in Chapter 2.3.3, the error is determined by the variation of drain-source and bulk leakage currents of M0 when it is OFF. To reduce the variation in drain-source leakage current, a bulk-source-connected M0 has been implemented. Additionally, to reduce the bulk leakage current of M0, M2 is ON to set the source voltage of M0 to V_{BE_rep}. In the meantime, M4 and M5 are ON to discharge V_{DAC} to the ground.

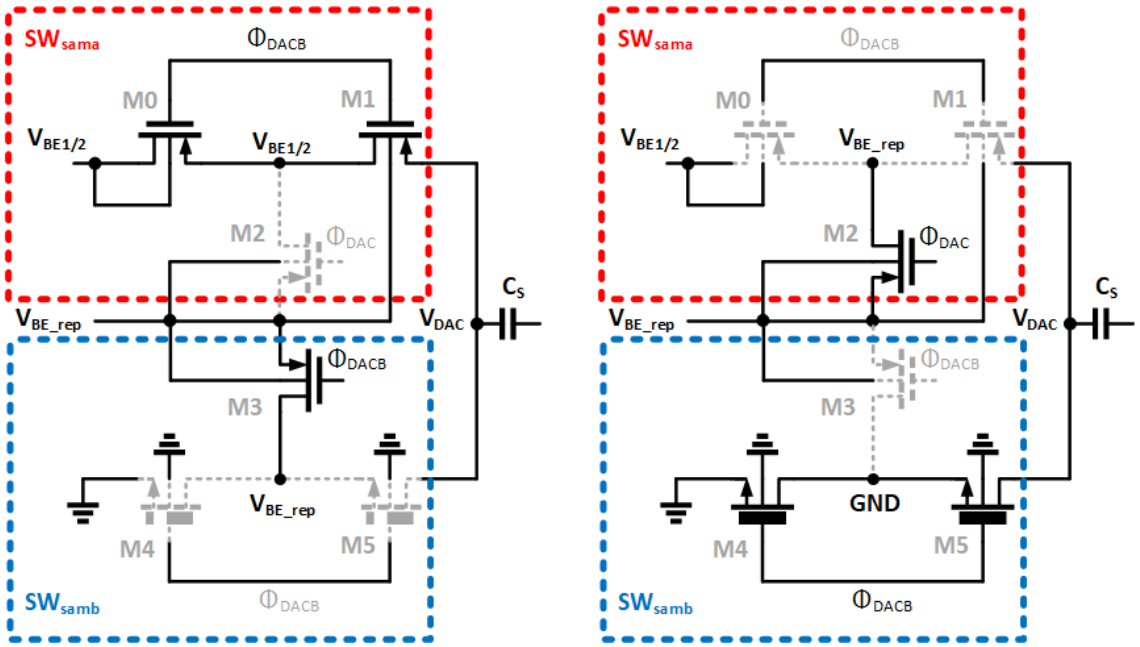


Figure 3.12 Two phases of a unit DAC element: $\Phi_{DAC} = 1$ (left), $\Phi_{DAC} = 0$ (right).

M0 and M1 are 0.88µm/0.18µm RVT PMOS devices to achieve fast settling when Φ_{DAC} is low. M2 and M3 are 0.22µm/0.18µm RVT PMOS to minimize the bulk leakage current. Additionally, M5 is 0.22µm/0.35µm HVT NMOS, and M4 is 0.22µm/1µm HVT NMOS, reducing the drain-source leakage currents. Consequently, each modified T-switch can drive a 40fF sampling capacitor with less than 60fA variation in leakage current, translating to ±0.05°C error.

3.2 Tracking Delta-Sigma Modulator

Figure 3.13 shows the block diagram of the tracking DT-DSM. The outputs of the front-end, V_{BE} and ΔV_{BE} are applied to an SC-DAC and a discrete-time integrator. C_s is the unit sampling capacitor of the SC-DAC, and C_{INT} is the integration capacitor. The DSM operates by balancing V_{BE} with $K\Delta V_{BE}$ (where K is the average of K_1 and K_2). The output of the first integrator V_{INT}(i) is quantized by a 1-bit quantizer. A feedforward DLF is implemented to transform the 1-bit bitstream output into a 6-bit D_{OUT}. To keep the DSM loop stable, the feedforward factor α is set to 0.125, while the second stage gain β is set to 0.25. K_1 and K_2 are decoded thermometer codes that control the SC-DAC.

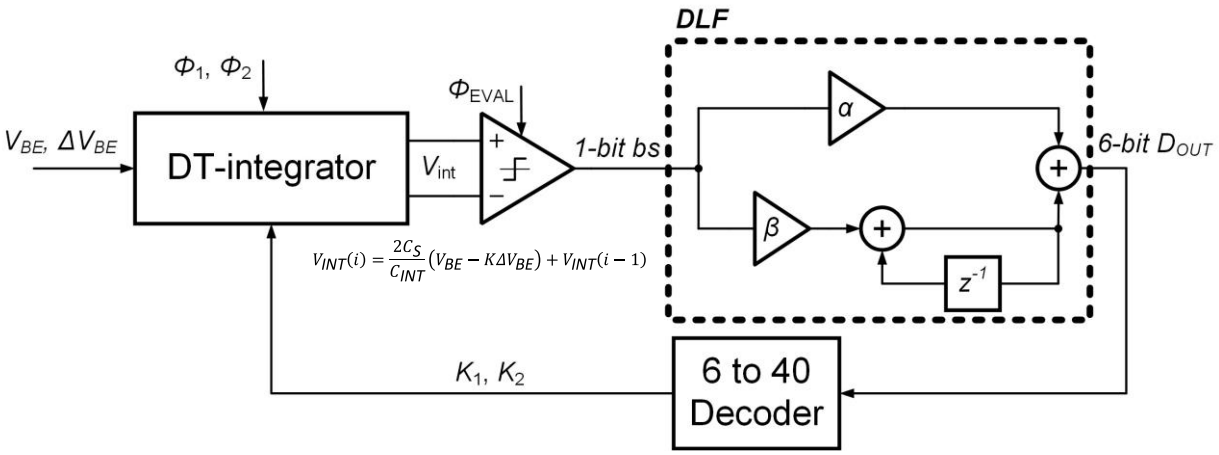


Figure 3.13 Block Diagram of the tracking DT-DSM.

3.2.1 Implementation of the first integrator

The block diagram of the first integrator is shown in Figure 3.14 (left), where C_{AZ} is the auto-zero capacitor. As mentioned earlier, 20 sampling capacitors are connected to the sampling switches separately for each half of the SC-DAC. These switches are controlled by different thermometer codes, K_1 and K_2, while Φ_1 and Φ_2 represent the sampling and integrating phases, respectively. The control signals $\Phi_{DAC1\text{-}40}$ are generated from the control signals Φ_{1D} and Φ_{2D}. For low-power designs, an energy-efficient inverter-based amplifier is employed as the first integrator [18], which stacks only 4 transistors in one branch, improving the output swing presented in [13].

As shown in Figure 3.14 (right), the control signals Φ_1 and Φ_2 controlling the integrator switches need to be earlier than Φ_{1D} and Φ_{2D} to avoid the charge injection caused by the switching-OFF of the sampling switches by $\Phi_{DAC1\text{-}40}$. Moreover, the evaluation signal (Φ_{EVAL}) for the comparator [12] arrives earlier than Φ_2. The signals Φ_{EVAL}, Φ_1, Φ_2, Φ_{1D}, and Φ_{2D} are generated by a non-overlapping clock generator.

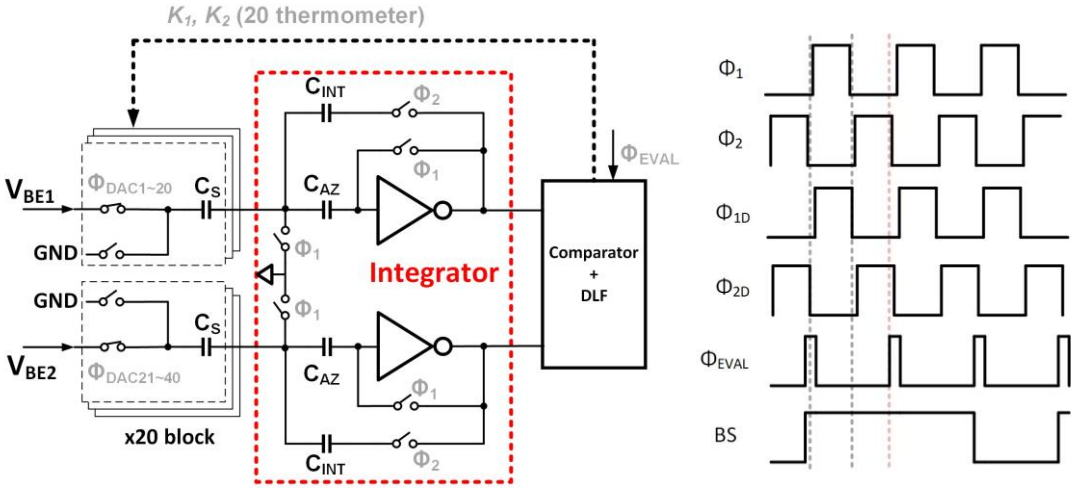


Figure 3.14 Block diagram of the first DT integrator (left) and time diagram (right).

Figure 3.15 illustrates the schematic of the inverter-based amplifier as the integrator. The total current of the amplifier is 20nA with a transconductance of 400nS to limit settling error in V_{BE} and ΔV_{BE} less than 0.01°C. To mitigate the offset and flicker noise, four auto-zero capacitors (C_{AZ}) of 1.6pF are used in series with the gates of the input transistors. However, the implementation of C_{AZ} introduces dynamic errors, as will be analyzed later.

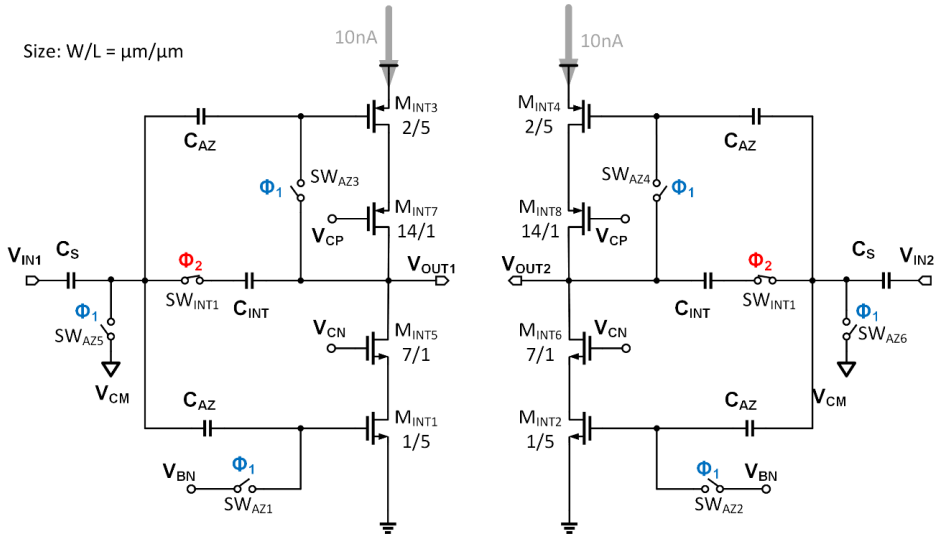


Figure 3.15 Schematic of the inverter-based amplifier.

The loop gain of the DSM depends on the gain of the first integrator. To achieve SQNR = 84dB, with a zoom-in factor of 24, the required open-loop gain of the first integrator ($A_{OL,INT}$) is calculated as below:

$$A_{OL,INT} > 10 \times \frac{SQNR}{24} = 78dB \tag{3.1}$$

To suppress amplifier nonlinearities, this integrator has achieved an open-loop gain larger than 86dB across corner and temperature with all transistors working in the saturation region. The maximum output swing of this inverter-based amplifier is ±200mV with 1V supply.

The single-ended operation of the first integrator is depicted in Figure 3.16. In Φ_1, the input of C_{INT} is disconnected, and the amplifier is auto-zeroing. At the end of Φ_1, the input voltage is sampled on C_S. In the following Φ_2, the auto-zero capacitors are disconnected from the output and connected in series with the input transistors when the amplifier integrates the sampled charge on C_{INT}.

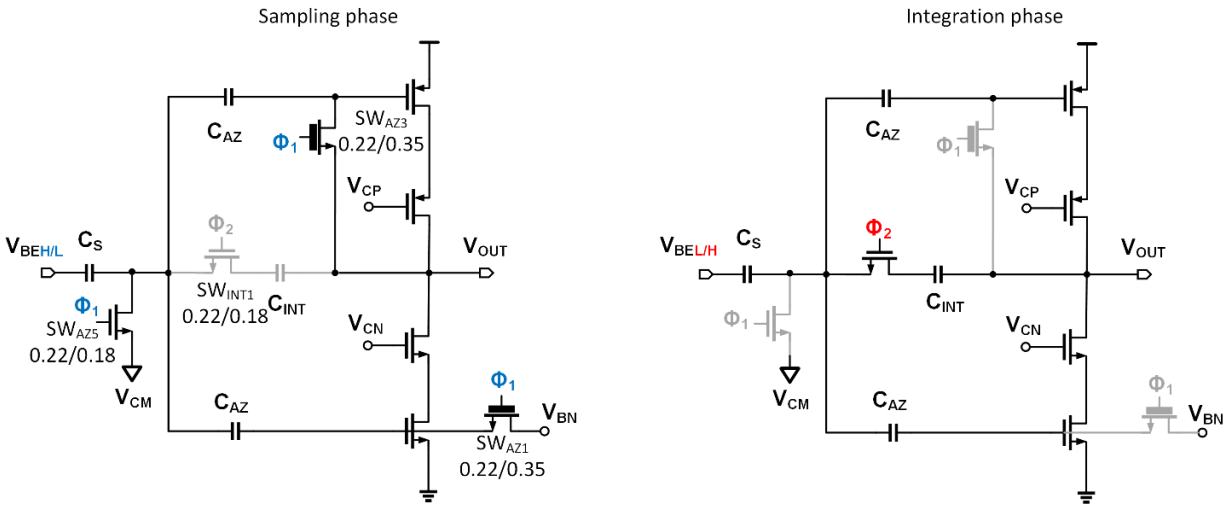


Figure 3.16 Operation of a single-sided inverter-based amplifier.

During the transition from Φ_1 to Φ_2, the charge injection of auto-zero switches SW_{AZ1-4} introduces common mode shift in the output common-mode voltage, which will be converted to a differential residue offset error on $V_{OUT,DIFF}$ due to mismatch in C_{AZ}. Similarly, during Φ_2, leakage current through C_{AZ} will also cause a common-mode drift, thus resulting in a residue differential offset error due to mismatch. Thus, the auto-zero switches SW_{AZ1-4} are chosen to be HVT NMOSs, sized as 0.22µm/0.35µm, to minimize the leakage current. An extra bootstrapped gate driving voltage $2V_{DD}$ is applied to these HVT switches to reach low ON-resistance, which is generated from charge pumps [19] and level shifters.

During the transition from Φ_2 to Φ_1, a charge injection caused by the switching-OFF of SW_{INT1} and SW_{INT2} leads to a common-mode drift on V_{INT}. During Φ_1, leakage current flowing through SW_{INT1} and SW_{INT2} results in an additional common-mode drift. This error is mitigated by using common-mode feedback (CMFB) in Φ_2. A scaled-down cross-coupling capacitive input CMFB, as implemented in [18], is employed to fix the input and output common-mode voltage of this amplifier. The amplitude of output common-mode drift is smaller than 5mV. The mismatch between C_{INT} introduces residue differential offset voltage on V_{INT}, which is mitigated by the system-level chopping.

A metal-insulator-metal (MIM) capacitor is chosen for C_S and C_{INT} because of its stable capacitance across process corners and temperature. The largest output swing of the first integrator is 100mV which is smaller than the maximum output swing of 200mV in this amplifier.

3.2.2 Implementation of the digital controller

The digital controller, based on the previous design [13], has been used to generate all the control signals for DMFE control, DSM control, DEM and Φ_{CHL}, as shown in Table 3.1.

Table 3.1 Input and output signals of the digital controller.

Signal name	Signal type	Signal state	Description
BS_IN(B)	Input	0/1	Quantizer positive (negative) output
Φ_1	Input	0/1	Non-overlapping clocks
Φ_2	Input	0/1	
Φ_{1D}	Input	0/1	Delayed non-overlapping clocks
Φ_{2D}	Input	0/1	
Φ_{AFE_OTA}	Output	0/1	(The bias amplifier) 1: enable, 0: disable
Φ_{AFE}	Output	0/1	(DMFE) 1: pre-charge phase, 2: conversion phase
Φ_{AFED}	Output	0/1	Delayed signal of AFE
AFE_CMDEM <7:0> Φ_{DEM}	Output	8-b thermometer	(CM connection) 1: to V_{BE1} 0: to V_{BE2}
DAC <39:0> Φ_{DAC}	Output	40-b thermometer	(C_S connection) 1: to V_{BE} 0: to ground
RST_CINT	Output	0/1	1: C_{INT} reset; 0: nothing
K<5:0>	Output	6-b binary	Digital outputs
State<1:0>	Output	2-b binary	0: bias phase Φ_A, 1: initial phase Φ_B, 2: tracking phase Φ_C, 3: DSM conversion phase Φ_D

Several new signals, Φ_{AFE_OTA}, Φ_{AFE}, and Φ_{AFED}, are assigned to control the DMFE depicted in Figure 3.17. During State 0 to 1 transition, the principal idea is to protect the most sensitive voltage V_{bias} by disconnecting the C_{bias} from the bias amplifier first. Then, R_{bias} and R_{beta} are shorted when the bias amplifier inputs are disconnected with the control of Φ_{AFED} to mitigate the influence caused by the parasitic capacitors. After one clock cycle, the bias amplifier is disabled by cutting the power supply. During State 2 to 3 transition, the DEM techniques are enabled for the front-end operating in the conversion phase.

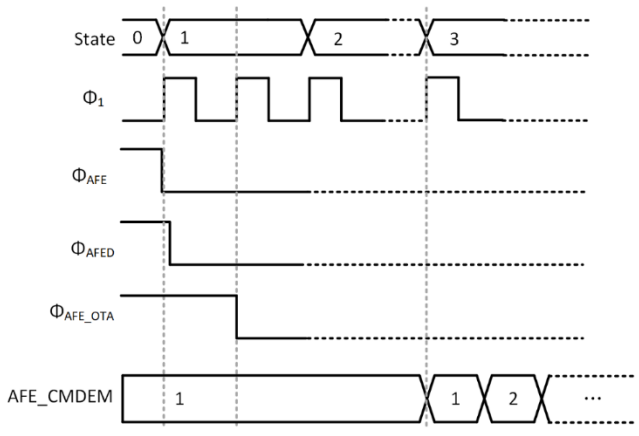


Figure 3.17 Time diagram of output signals controlling the DMFE.

3.3 Power breakdown

A summary of power consumption for both the analog and digital circuitry is listed in Table 3.2 from post-layout simulations. The total power consumption is 80nW. The DMFE and the tracking DSM consume 23.75% and 25% of the total power, respectively, while the remaining 50% of the power is consumed by the digital controller and clock generator. This indicates that further power savings can be made by simplifying the digital circuitry.

Table 3.2 Power breakdown of analog and digital blocks this work.

Blocks	Sub-circuit	Power (1V supply)	Power contribution (%)
DMFE	AFE core	12nW	15
	Bias amplifier	3nW	3.75
	Additional biasing branch	4nW	5
Tracking DSM	The first integrator	20nW	25
	Comparator	1nW	1.25
Clock generator	Voltage doubler	3nW	3.75
	Non-overlapping clock	7nW	8.75
Digital controller		30nW	37.5
Full circuit		80nW	100

3.4 Summary

In this chapter, the circuit-level implementations of the front-end, readout are discussed in detail. A power breakdown from post-layout simulations is presented. The next chapter will present the measurement results of 18 prototype chips.

4 Measurement results

This chapter describes the measurement setup and the measurement results of 18 prototype chips fabricated in a TSMC 180nm CMOS technology. The die micrograph is shown in Figure 4.1. The area of the prototype chip is 0.228 mm^2.

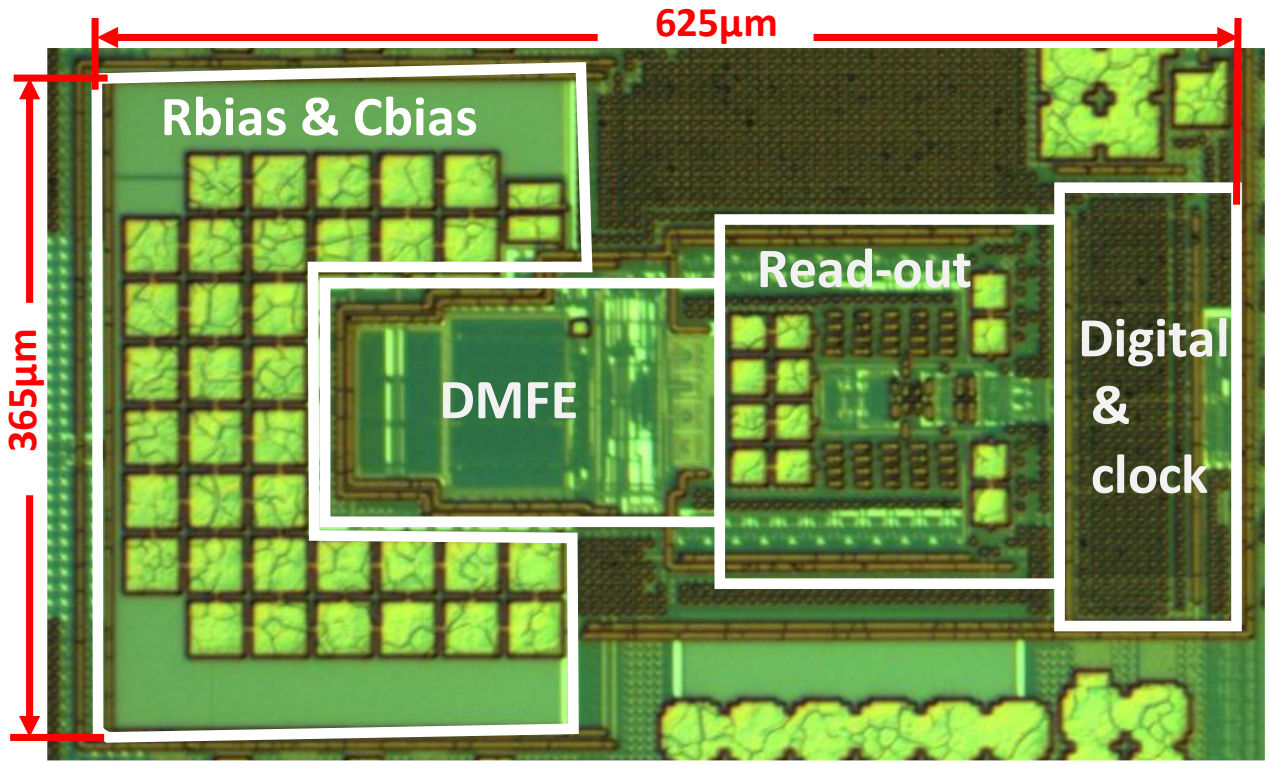


Figure 4.1 Die micrograph.

4.1 Measurement setup

The block diagram of the measurement setup is illustrated in Figure 4.2. The prototype chips were placed in a Votsch 7004 oven, whose temperature was swept from -40°C to 85°C. The chips are mounted on a PCB board and are placed inside a large metal box which acts as a low-pass thermal filter. A layer of thermal paste ensures close contact between the top surfaces of the chips and the metal surface to ensure good thermal conductivity. The temperature of the block is monitored by a Kelvin-connected Pt-100 sensor, which is calibrated to an inaccuracy of less than 0.01°C. Once the temperature in the block settles, the temperature acquired from the Pt-100 will be equal to the temperature of the prototype chips. The Pt-100 is read out by an Keithley 2001A 8.5-bit multimeter. The master clock is generated by the Keysight 33120A function generator. The control bits are written to the on-chip shift registers generated from FPGA. The power supply, function generator, multimeter, and oven are all controlled by the PC through GPIO ports.

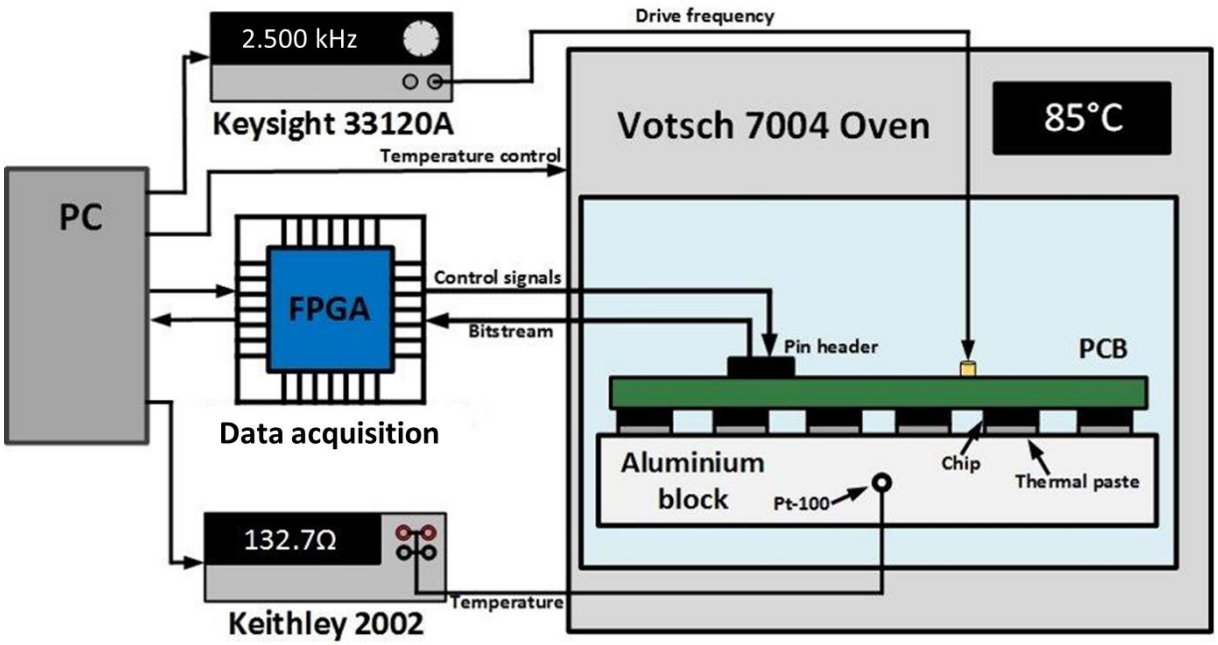


Figure 4.2 Block diagram of measurement setup.

4.2 Power breakdown

The measurement results of the power breakdown are shown in Table 4.1. The measured power consumption is 130nW, which is 50nW larger than the simulated result. The power can be reduced by disabling the β-compensation, which reduces I_{bias}.

In this design, the digital controller is synthesized by Cadence Genus Synthesis Solution with the standard library provided by TSMC. However, these designs are optimized for a 1.8V supply, which is not suitable for <1.1V supply voltages, especially at low temperatures. Moreover, the digital pads, usually powered by a 1.8V supply, were accidentally connected to the on-chip digital supply instead of the padring power supply. To make the pads operational from -40°C to 85°C, the digital supply has to be increased to 1.3V.

The power consumption of the digital logic is 58nW, which is twice as much as the post-simulation results. This is caused by a larger digital supply and inaccurate simulation models. The power consumption of the analog part is 72nW, which is 25% higher than the simulated results.

Table 4.1 Power analysis

Block		Supply voltage	Supply current	Power consumption	Power contribution
Analog	β-comp	1.2V	60nA	72nW	55%
	no β-comp		45nA	54nW	
Digital		1.3V	45nA	58nW	45%
Total	β-comp	1.2/1.3V		130nW	100%
	no β-comp			113nW	

4.3 Accuracy

Figure 4.3 shows the untrimmed decimated output values, X, for 18 chips from –40°C to 85°C. The expected non-linear behavior is observed.

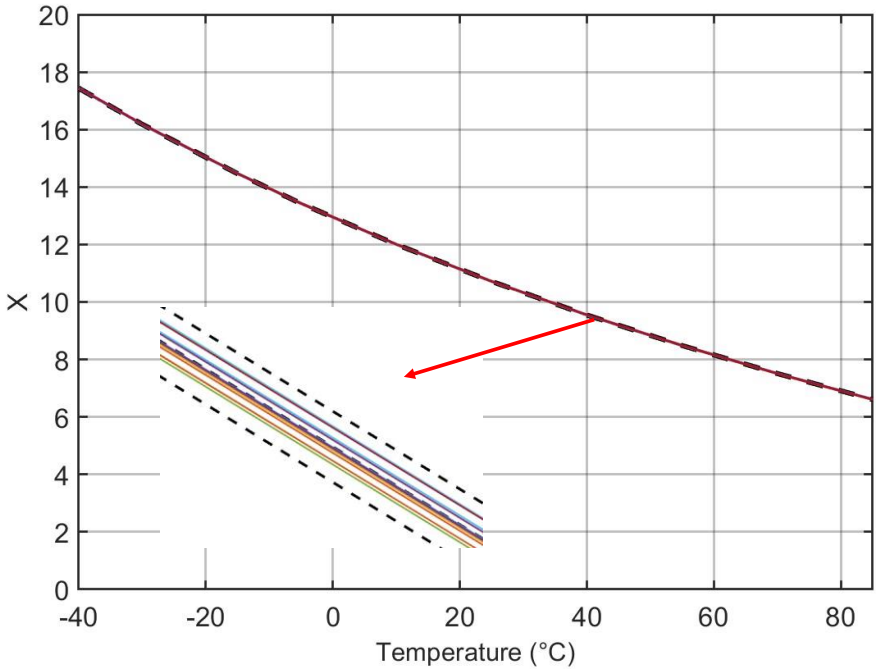


Figure 4.3 Measured output of the sensors from –40°C to 85°C.

The decimated output X is converted to temperature using Equations 1.4 and 2.12. The coefficients are α =14.5, A = 536.98, B = -283.82. The residual non-linearity of the 18 measured sensors is above 0.1°C at high temperatures without trimming, and above 0.05°C with trimming, as shown in Figure 4.4. This is caused by the increase of leakage currents of switches.

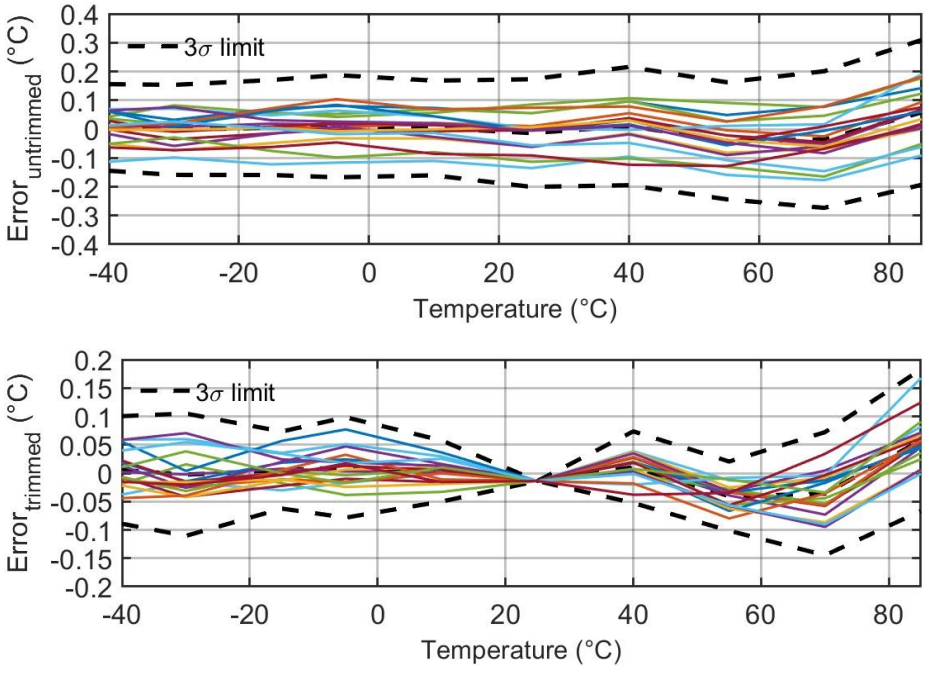


Figure 4.4 Temperature error of 18 sensors with Equation 2.12: untrimmed (top), trimmed (bottom)

A high-order (5th) polynomial correction is done to remove the systematic non-linearity, as will be shown below. Figure 4.5 shows the untrimmed temperature error of 18 chips after linearization with β-compensation. As expected, the temperature error is PTAT, due to the spread in V_{BE}, resulting in a 3σ inaccuracy of ±0.25°C.

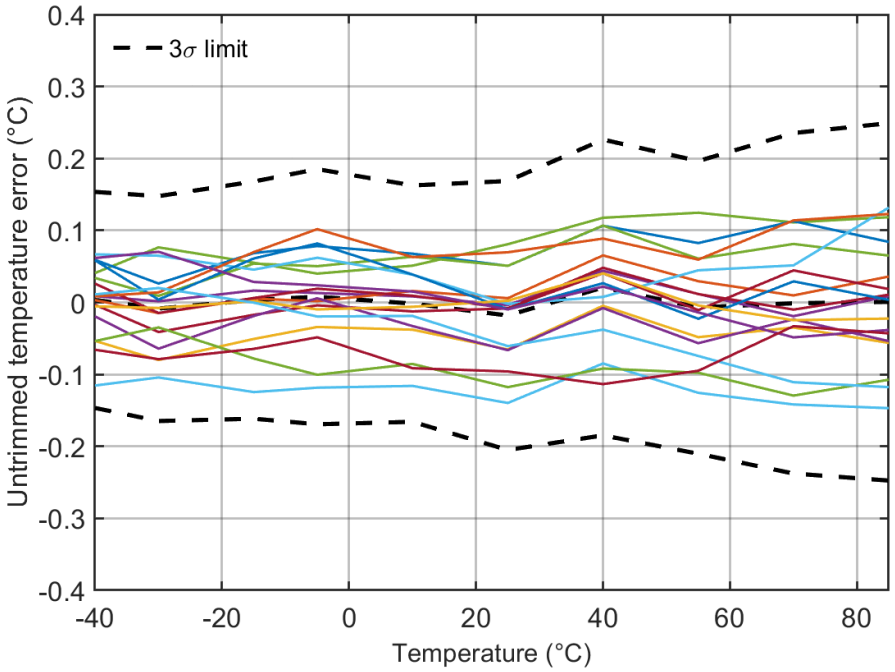


Figure 4.5 Untrimmed temperature error of 18 sensors (β_{comp}).

This PTAT error is corrected by applying an offset trim to X [17]. Figure 4.6 shows the temperature error of 18 sensors after a 1-point trim. The inaccuracy from -45°C to 85°C is improved to ±0.13°C (3σ).

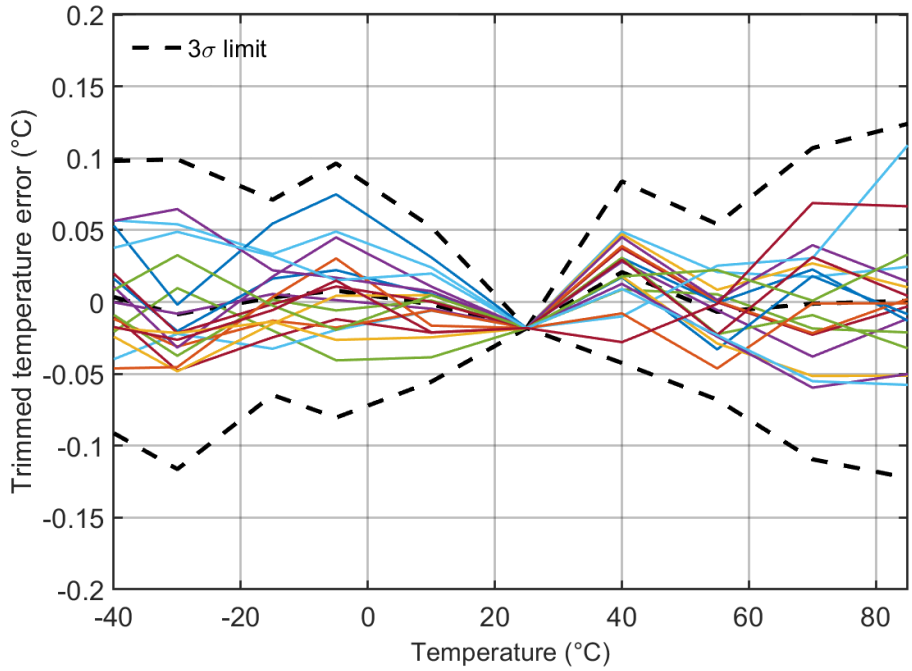


Figure 4.6 1-point trimmed temperature error of 18 sensors (β-comp).

Another measured temperature error for the chips without β-compensation after 1-point calibration is shown in Figure 4.7. The inaccuracy from -45°C to 85°C is slightly larger than ±0.15°C (3σ). This also indicates that β-compensation only provides limited improvement at low temperatures.

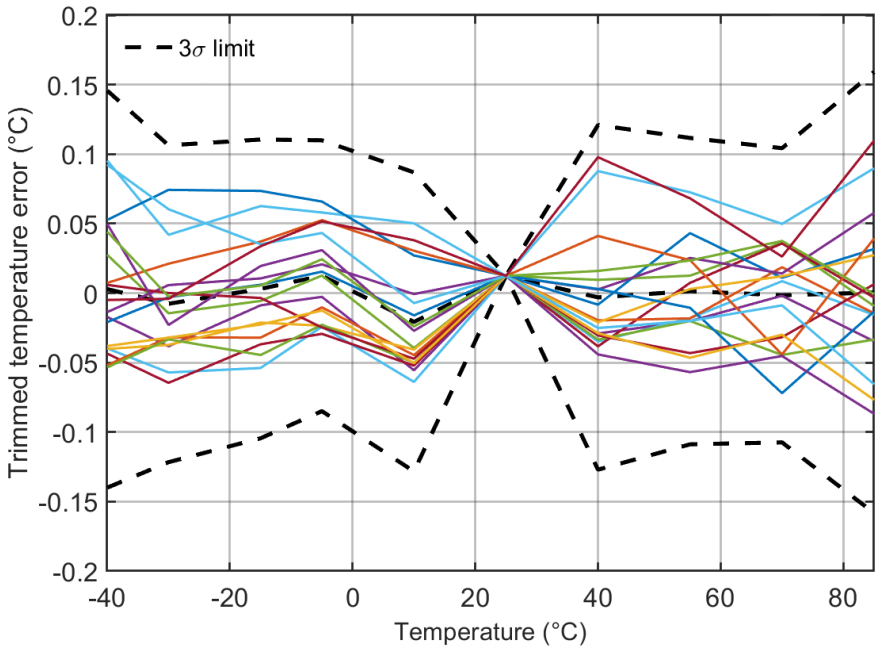


Figure 4.7 1-point trimmed temperature error of 18 sensors (no β-comp).

4.4 Resolution

The measured noise spectrum of the output bitstream at room temperature (20°C) is shown in Figure 4.8. To plot the FFT, the bitstream outputs of the conversion phase are stitched together. The bitstream output contains 300 conversion cycles. For each conversion cycle, the conversion time (T_{conv}) is set to 80ms, while the settling time is set to 20ms.

All dynamic techniques in this design have been applied during the measurement. Φ_{CHL} tones are visible in the spectrum at the multiples of frequency $1/T_{conv}$ = 12.5Hz, caused by the large systematic error and mismatch presented in the sensor. A first-order noise shaping can be seen at high frequencies. The use of barrel-shifting DEM increases the noise-floor by only 2dB. Better results could be obtained with bitstream-controlled DEM [16].

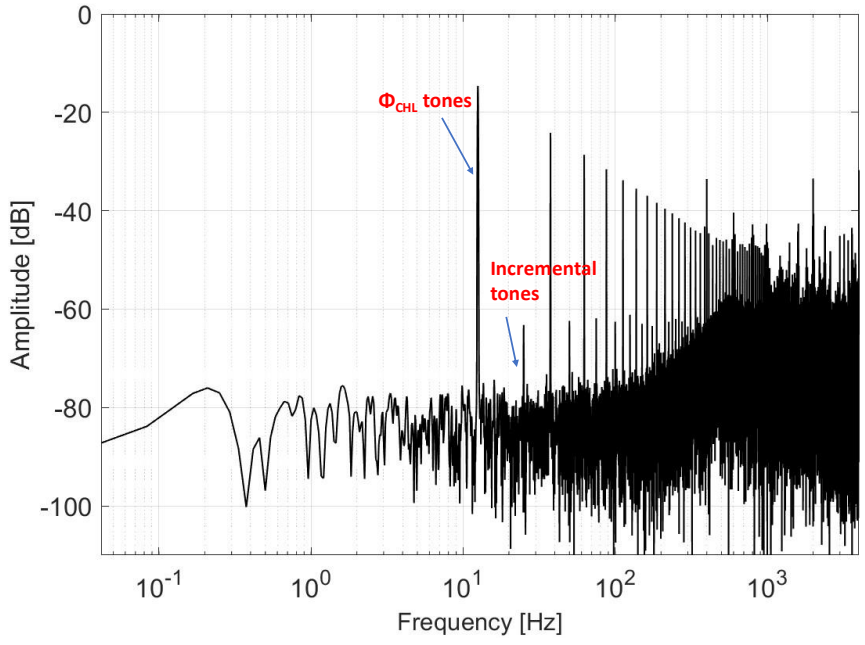


Figure 4.8 Noise spectrum of a prototype sensor.

Figure 4.9 shows the measured resolution and Figure of Merit (FoM) versus the conversion time, averaged for 18 chips. When the conversion time is longer than 80ms, corresponding to 200 DSM cycles (OSR = 100), the resolution is limited by thermal noise, and the best FoM is achieved. With a conversion time of 100ms, the resolution is 32mK, and the FoM is 13 pJ°C^2. The maximum possible conversion time is limited to 250ms by the design of the digital controller.

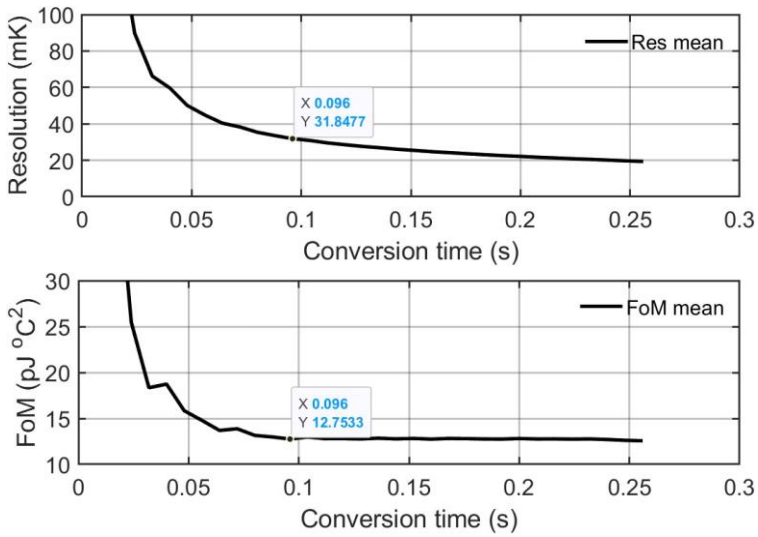


Figure 4.9 Resolution (top) and the FoM over the conversion time (bottom).

4.5 Power supply sensitivity

For supply voltages ranging from 1.1V to 1.3V, the measured Power Supply Sensitivity (PSS) of the sensors at room temperature (27°C) is shown in Figure 4.10. A PSS of 0.1°C/V was obtained. Since a voltage doubler is used in the design, the maximum power supply voltage is limited to 1.3V to prevent the drain-source voltage from exceeding the breakdown value of 1.8V NMOS.

As the supply voltage decreases, so does the voltage headroom for the current source, resulting in large temperature errors at supply voltages below 1.1V. Similarly, as the supply voltage increases, the cascode stage of the front-end moves closer to the triode region, leading to temperature errors.

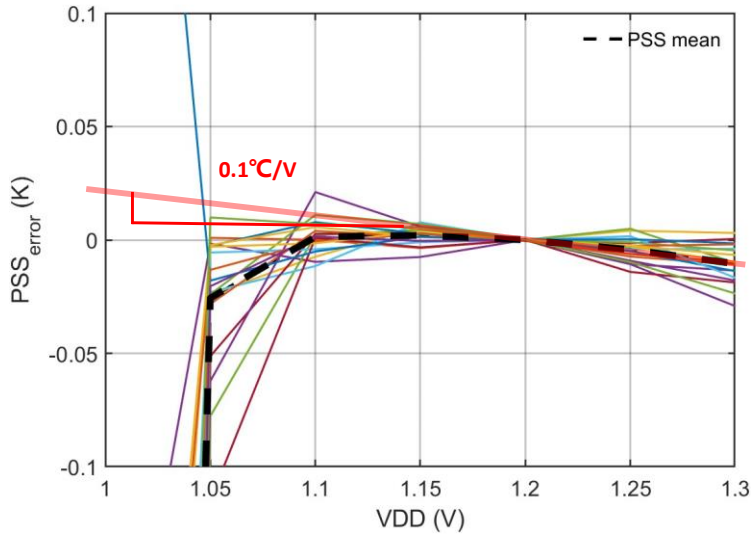


Figure 4.10 Power supply sensitivity for 18 sensors at 27°C

4.6 Summary

In Table 4.2, the performance of the BJT temperature sensor is summarized and compared with other low-power designs. The sensor achieves an accuracy (0.15°C (3σ) from -45°C to 85°C) which is comparable with the other designs, while dissipating the lowest power (130nW).

Table 4.2 Performance summary and comparison to the state-of-the-art.

	VLSI'22 [13]	ISSCC'14 [9]	ISSCC'23 [12]	**This work**
Core type	NPN	DT-MOS	PNP	**PNP**
Technology	180nm BCD	160nm	180nm	**180nm**
Power [nW]	210	600	810	**130**
Supply voltage [V]	1.25	0.85	0.95	**1.2/1.3**
Temperature range [°C]	-15 to 85	-40 to 125	-55 to 125	**-40 to 85**
3σ inaccuracy [°C] (trim point)	±0.4 (0) ±0.15 (1)	±1 (0) ±0.4 (1)	±0.45 (0) ±0.15 (1)	**±0.25 (0)** **±0.15 (1)**
Relative inaccuracy [%] (trim point)	0.80 (0) 0.30 (1)	1.2 (0) 0.50 (1)	0.50 (0) 0.17 (1)	**0.40 (0)** **0.24 (1)**
Conversion time [ms]	50	6	128	**100**
Resolution (mK)	15	63	1.8	**32**
Area (mm^2)	0.058	0.085	0.25	**0.228**
Res. FoM ($pJ°C^2$)	2.4	14	0.34	**13**

5 Conclusion and future work

5.1 Conclusion

A 130nW BJT-based temperature sensor with a dual-mode front-end has been designed and fabricated in TSMC 180nm CMOS technology. A novel PNP-based ultra-low power dual-mode front-end is proposed with a unit bias current of 1nA, which reduces the power consumption of both the front-end and readout circuits. A tracking DSM is used to reduce the conversion time and relax the driving requirements from the front-end. With dynamic techniques, non-idealities such as component mismatch induced errors can be reduced significantly. Moreover, measurements show that this sensor achieves ±0.15°C inaccuracy from -40°C to 85°C (RIA = 0.24%) with only 1-point calibration while consuming lesser power than other state-of-the-art low-power temperature sensors, as shown in Figure 5.1. This work indicates that sub-100nW BJT-based temperature sensors can probably still achieve high accuracy, which is attractive for all low-power applications.

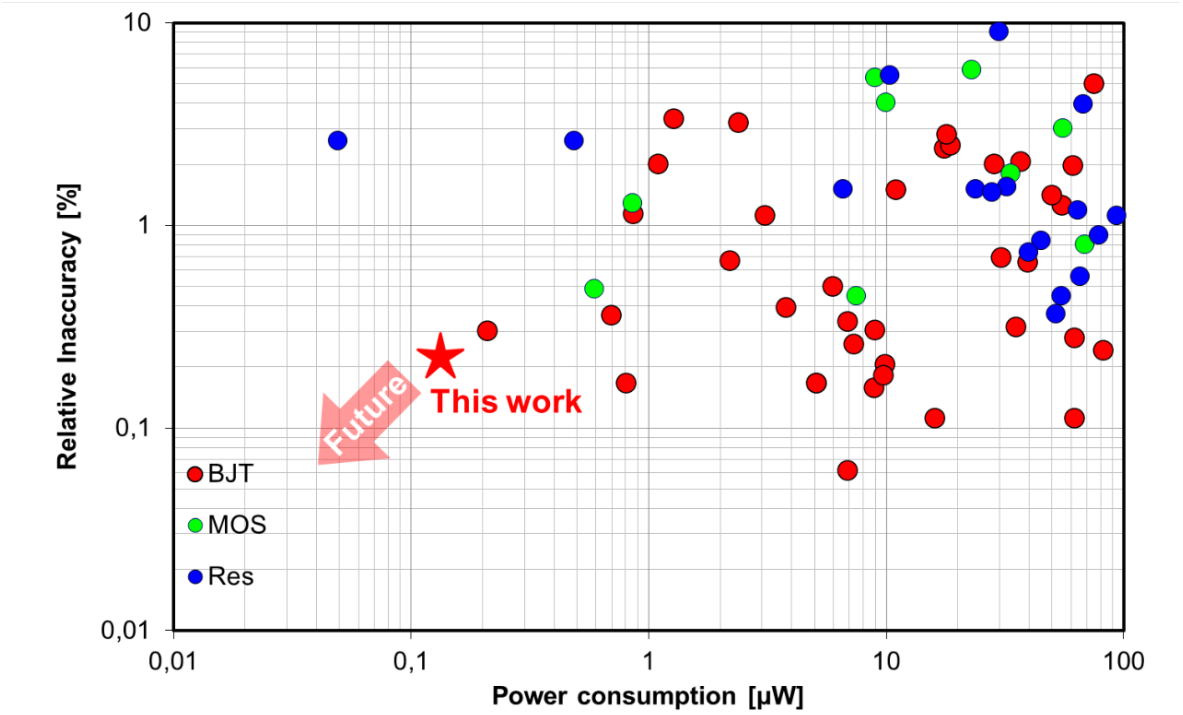


Figure 5.1 Achieved specification among the state-of-the-art low-power temperature sensors.

5.2 Future work

5.2.1 Reducing power

5.2.1.1 Lowering supply voltage

The total power consumption can be reduced by lowering the supply voltage. For the digital supply, main limitation to lower the supply voltage to below 1V is the setting of Synopsys-design-constraints (.sdc) with the automatic digital synthesizer. With a manual design in the digital circuit, the standard digital library can achieve 0.85V, when clocked at 1MHz [12].

To reduce the analog supply, the headroom of the BJT transistors is the main limitation. The measured PSS shows that the error increases significantly when the analog supply voltage is lower than 1.1V. Thus, the headroom of the current sources needs to be decreased by at least 50mV to achieve sub-1V operation. This can be achieved by operating them in moderate inversion, at the expense of larger area and mismatch.

5.2.1.2 Lowering bias current

As seen from the measurement results, significant error is observed at low temperatures, which is probably due to the PNP's small and non-linear β, especially at low current levels. The β-compensation technique provides only limited help, because it does not compensate for the spread in $\Delta\beta$. To mitigate this, increasing the effective β is the only way, which can be achieved by adding an auxiliary transistor current source to compensate for the base current [4]. Moreover, in the 180nm BCD process, a high β PNP is available, which should have significantly lower β-related errors.

The variation in leakage current needs to be considered when reducing the bias current. For the sampling switches, when Φ_{DAC} is high, the drain-source leakage varies since the input voltage V_{BE} varies with temperature. To fix this problem, as shown in Figure 5.2, a bulk-source connection could be applied to M5 by placing it in a deep Nwell. The leakage current between the deep Nwell and the body will not be an issue when the operating temperature is lower than 100℃. From -40℃ to 85℃, simulation results show that the variations in the leakage current are reduced from 700fA to 100fA.

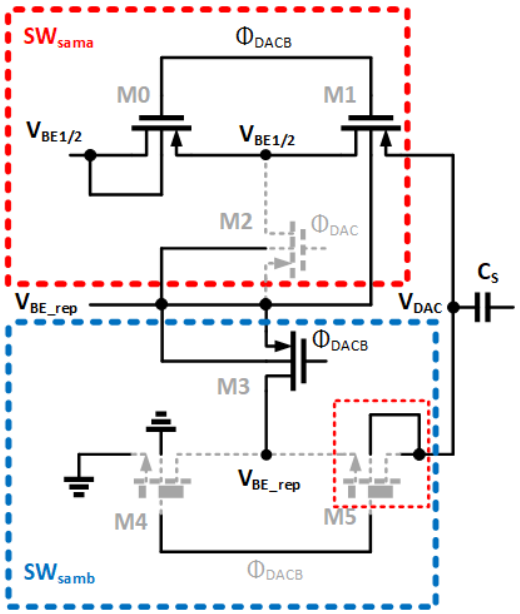


Figure 5.2 Schematic of a new sampling switch group when $\Phi_{DAC} = 1$.

5.2.1.3 Simplifying digital controller

Saving the power of the digital block is feasible. Besides changing the type of transistors (LVT) in the standard digital library, the complexity of the digital controller can be improved by reducing the power of C_S DEM and DLF. In this design, the tracking DSM can achieve a smaller reference voltage ($0.5\Delta V_{BE}$) than that in conventional zoom ADCs ($2\Delta V_{BE}$ [11]), which doesn't require over-ranging and has a simpler digital controller (only one DLF). However, a typical IIR filter is applied as the DLF [15], which contains such disadvantages: an internal accumulator and becoming instable in some case. To remove the power-hungry accumulator and keep the loop always stable, the IIR filter can be replaced by a FIR filter, whose output has no relationship with its last output result. Furthermore, the tracking controller can be replaced by a SAR logic in the coarse conversion, as Φ_C in

Figure 2.10. As shown in Figure 5.3, the DLF is replaced by two FIR-DAC, one drives the feedback loop, and another drives the compensation loop. The reference of this FIR-DAC is $2\Delta V_{BE}$, translating into $0.5\Delta V_{BE}$ step at the input of the integrator, as the expense of larger OSR to achieve the same SQNR. The zoom-in ADC cover the temperature range from -55 to 125 with a reference from $4\Delta V_{BE}$ to $20\Delta V_{BE}$, and it has a step of $2\Delta V_{BE}$. Due to the reduced complexity of this digital filter and the intrinsic DEM characteristics in FIR DAC, the power can be reduced further.

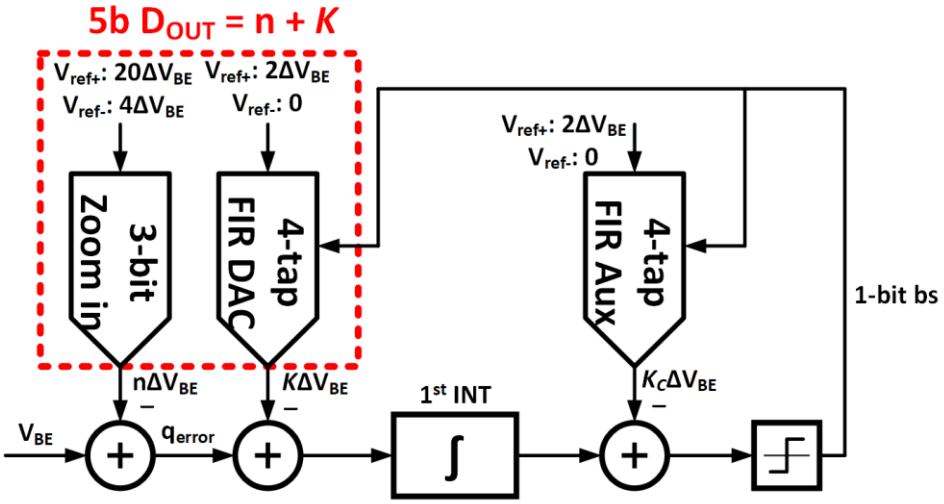


Figure 5.3 Readout topology of FIR-DAC with SAR logic.

The timing diagrams of these two readout architectures are shown in Figure 5.4. The new FIR DSM reduces the time of the tracking phase, while keeping the advantage of tracking DSM in the DSM conversion phase, which can mitigate the over-ranging issue caused by the sample noise. The output voltage range, as shown in Figure 5.5, of the first integrator in the new FIR DSM is smaller than 100mV, which is similar to the tracking DSM.

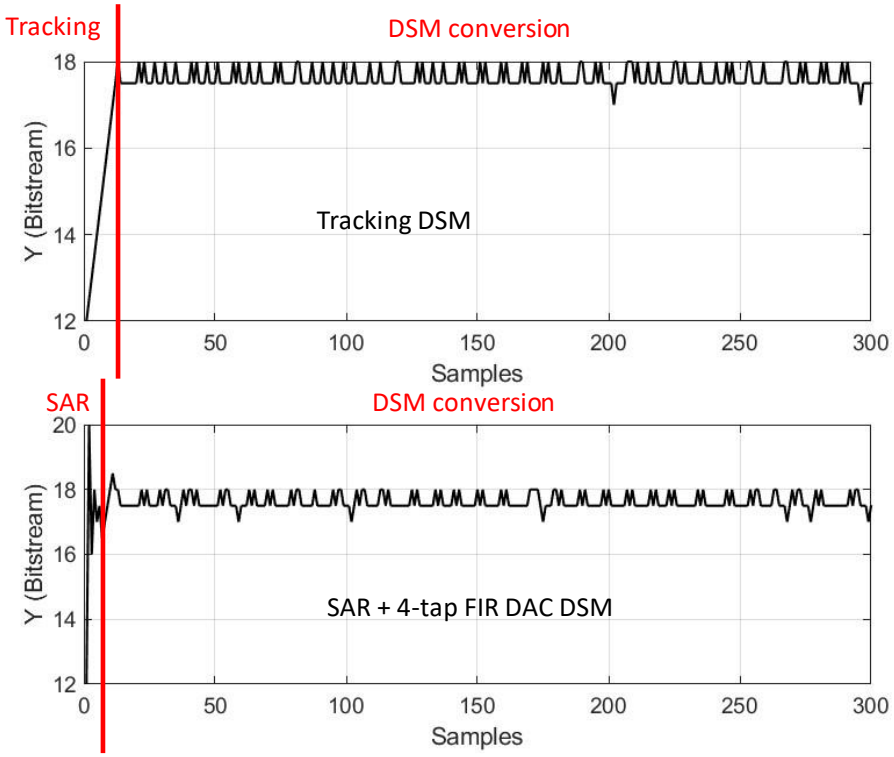


Figure 5.4 Timing diagrams of two readout architectures (simulated with noise in MATLAB).

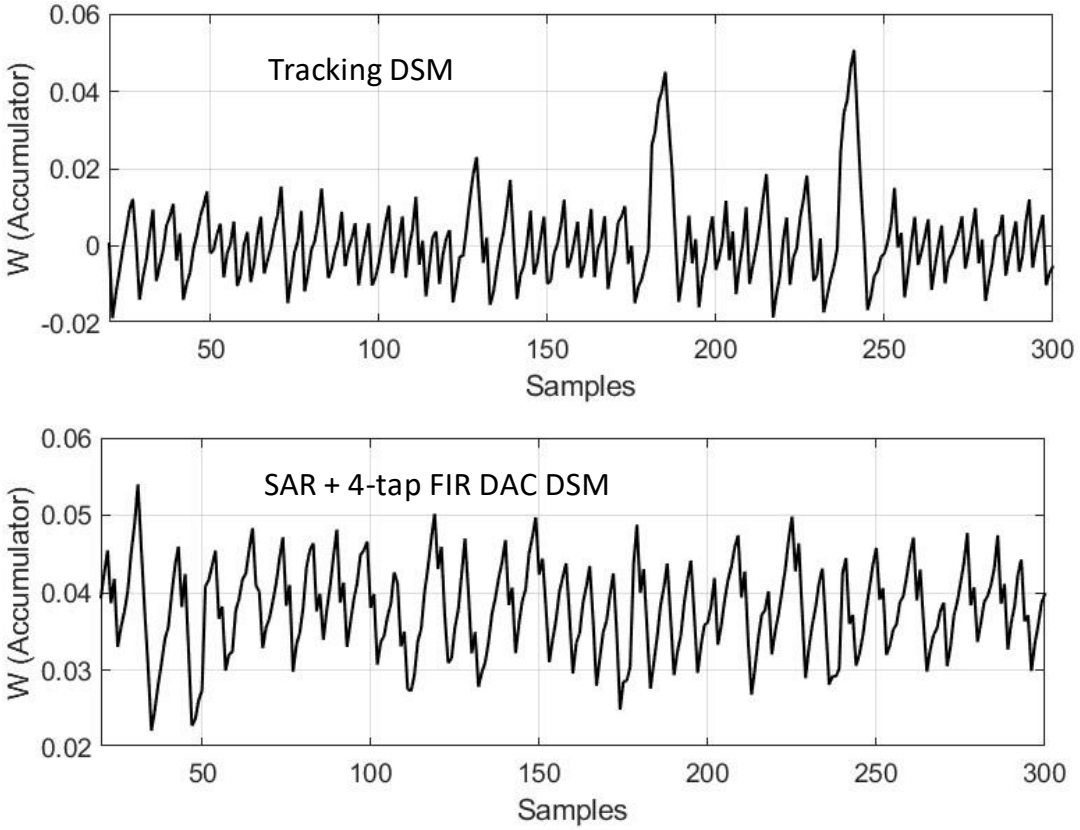


Figure 5.5 Integrator differential output voltage of two readout architectures.

As shown in Figure 5.6, to achieve the same SQNR as the tracking DSM (OSR = 100), the OSR of proposed readout architecture needs to be increased to 200.

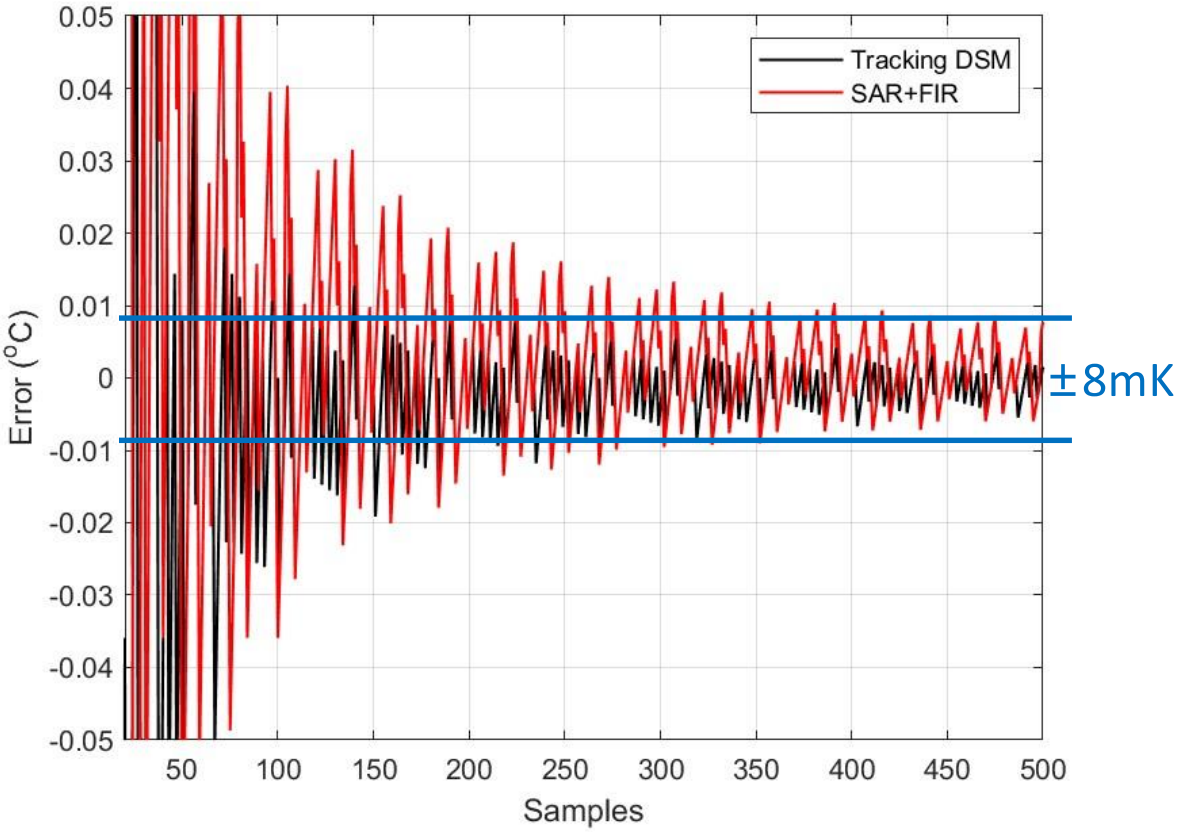


Figure 5.6 Error behavior of two readout architectures.

5.2.2 Reducing area

The area of temperature sensors is usually smaller than 0.1mm^2 [3], which is not achieved in this design. Thus, the area reduction must also be considered for the next design. In this design, the stacked bias resistor and capacitor occupy one-third of the area of the die, which is 0.07mm^2. To reduce the area of the resistor, a hipo resistor with higher sheet resistance needs to be used. The sheet resistance of the p-poly resistors in the 180nm BCD process is 5 times larger than the p-poly resistors in the standard 180nm process. While reducing the area of the bias capacitor (60pF), stability can be kept if the g_m of the bias amplifier decreases with the capacitance of C_{bias} in the same ratio.

Bibliography

[1] Analog Devices, “Lithium Coin-Cell Batteries: Predicting an application lifetime,” [Online]. Available: https://www.analog.com/en/technical-articles/lithium-coincell-batteries-predicting-an-application-lifetime.html

[2] Analog Devices, “Understanding Voltage-Reference Temperature Drift,” [Online]. Available: https://www.analog.com/en/technical-articles/understanding-voltagereference-temperature-drift.html

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

[4] M. Pertijs and J. Huijsing, Precision Temperature Sensors in CMOS technology, Analog Circuits and Signal Processing Springer, 2006.

[5] G. C. M. Meijer, Guijie Wang and F. Fruett, "Temperature sensors and voltage references implemented in CMOS technology," in IEEE Sensors Journal, vol. 1, no. 3, pp. 225-234, Oct 2001.

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

[7] G. C. Meijer, Integrated circuits and components for bandgap references and temperature transducers, PhD Thesis: TU Delft, 1982.

[8] 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.K2 at 150°C," 2022 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, USA, 2022, pp. 72-74.

[9] K. Souri, Y. Chae and K. Makinwa, "A CMOS temperature sensor with a voltage-calibrated inaccuracy of ±0.15°C (3σ) from −55 to 125°C," 2012 IEEE International Solid-State Circuits Conference, San Francisco, CA, USA, 2012, pp. 208-210.

[10] M. Terauchi, “Temperature dependence of the subthreshold characteristics of dynamic threshold MOSFETs and its application to an absolute-temperature sensing scheme for low-voltage operation,” Japanese Journal of Applied Physics, vol. 46, pp. 4102 - 4104, 2007.

[11] K. Souri, Y. Chae, F. Thus and K. Makinwa, "12.7 A 0.85V 600nW all-CMOS temperature sensor with an inaccuracy of ±0.4°C (3σ) from −40 to 125°C," 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), San Francisco, CA, USA, 2014, pp. 222-223.

[12] Z. Tang, S. Pan and K. A. A. Makinwa, "23.5 A Sub-1V 810nW Capacitively-Biased BJT-Based Temperature Sensor with an Inaccuracy of ±0.15°C (3σ) from −55°C to 125°C," 2023 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, USA, 2023, pp. 22-24.

[13] T. Someya, V. Van Hoek, J. Angevare, S. Pan and K. Makinwa, "A 210nW BJT-based Temperature Sensor with an Inaccuracy of ±0.15°C (3σ) from −15°C to 85°C," 2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), Honolulu, HI, USA, 2022, pp. 120-121.

[14] M. Motz, U. Ausserlechner, W. Scherr and B. Schaffer, "An Integrated Magnetic Sensor with Two Continuous-Time ΔΣ-Converters and Stress Compensation Capability," 2006 IEEE International Solid State Circuits Conference - Digest of Technical Papers, San Francisco, CA, USA, 2006, pp. 1151-1160.

[15] H. Son et al., "A Low-Power Wide Dynamic-Range Current Readout Circuit for Ion-Sensitive FET Sensors," in IEEE Transactions on Biomedical Circuits and Systems, vol. 11, no. 3, pp. 523-533, June 2017.

[16] M. A. Pertijs and J. H. Huijsing, "A sigma-delta modulator with bitstream-controlled dynamic element matching," Proceedings of the 30th European Solid-State Circuits Conference, Leuven, Belgium, 2004, pp. 187-190.

[17] 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," in IEEE Journal of Solid-State Circuits, vol. 52, no. 4, pp. 1044-1052, April 2017.

[18] Y. Chae and G. Han, "Low Voltage, Low Power, Inverter-Based Switched-Capacitor Delta-Sigma Modulator," in IEEE Journal of Solid-State Circuits, vol. 44, no. 2, pp. 458-472, Feb. 2009.

[19] J. -M. Baek, D. -J. Seo, J. -H. Chun and K. -W. Kwon, "A Dual Charge Pump for Quiescent Touch Sensor Power Supply," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 59, no. 11, pp. 780-784, Nov. 2012.
[20] S. Pavan, "Finite-impulse-response (FIR) feedback in continuous-time delta-sigma converters," 2018 IEEE Custom Integrated Circuits Conference (CICC), San Diego, CA, USA, 2018, pp. 1-8.