Compact Resistor-Based Temperature Sensors for On-Chip Thermal Monitoring

Angevare, J.

DOI
10.4233/uuid:d6a157f1-9e33-4f39-9efa-ac83723a9128

Publication date
2024

Document Version
Final published version

Citation (APA)
Angevare, J. (2024). *Compact Resistor-Based Temperature Sensors for On-Chip Thermal Monitoring*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:d6a157f1-9e33-4f39-9efa-ac83723a9128

Compact Resistor-Based Temperature Sensors for On-Chip Thermal Monitoring

Jan Angevare

Compact Resistor-Based Temperature Sensors for On-Chip Thermal Monitoring

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.dr.ir.
T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op Woensdag
15 Mei 2024 om 15:00 uur

door

Johannes Adrianus ANGEVARE

Ingenieur in de Micro-Elektronica,
Technische Universiteit Delft, Nederland,
geboren te Leiden, Nederland.

Dit proefschrift is goedgekeurd door de promotor.

Samenstelling promotiecommissie bestaat uit:

Rector Magnificus,	voorzitter
Prof. dr. K. A. A. Makinwa,	Technische Universiteit Delft, promotor

Onafhankelijke leden:

Dr. J. Shor,	Bar-Ilan U. Israël
Dr M.H. Perrott,	Texas Instruments, USA
Dr. Y. Chae,	Yonsei U. Sourth Korea
Prof.dr.ir. K. Bult,	TU Delft / Ethernovia, NL
Dr.ir. M.A.P. Pertijs	Technische Universiteit Delft
Prof.dr. L.C.N. de Vreede	Technische Universiteit Delft

Published and distributed by: Ipskamp Drukkers B.V.

ISBN 978-94-6473-504-8

Printed in the Netherlands.

Contents

Chapter 1 | Introduction

Modern System-on-Chips (SoCs) contain billions of transistors in a tiny area. So when they are operated at high clock frequencies, they can become very hot (>100°C), very quickly (in a few ms) [1]. Although advanced nodes require less switching power, their total switching activity tends to increase due to the increasing number of devices per unit area and/or higher switching frequencies. As a result, the effective energy density is increased, and self-heating is exacerbated. This self-heating can cause the chip to overheat, which in turn leads to logic errors, rapid aging, and in extreme cases, even to permanent damage [2].

1.1: Thermal Management

To prevent overheating, SoCs often require careful thermal design using heatsinks and cooling fans. Even with these measures in place, many of these devices can still overheat when left running at their maximum performance level. Therefore, throttling strategies are required that perform dynamic voltage and frequency scaling (DVFS) when the chip is close to overheating [3]. The costs of these solutions are justified by the large performance boost that is gained by running the SoCs at higher thermal design points (TDPs).

1.2: On-Chip Temperature Sensing

To detect the thermal state of the chip, temperature sensors are required. Since SoCs can exhibit significant local self-heating (hotspots) where the local temperature can easily be 10°C higher than that of the surrounding die [2], each of these potential hotspots must be monitored by a temperature sensor. Since many of these sensors will be placed close to dense digital circuitry, they should be small (<<10 000 μm^2) and tolerant of supply spikes and substrate noise. It is thus desirable for such sensors to operate from the digital supply voltage, and to maintain their required accuracy (~1°C) in the presence of supply voltage variations. This means that their resolution should also be commensurate, in the order of 0.1°C (rms). Furthermore, since thermal transients may have rise and fall times in the order of a few milliseconds, such sensors should have sampling rates of at least 1 kSa/s. However, their power consumption only needs to be a small fraction of the total system power, which is not difficult to achieve in most applications.

1.2.1 – BJT-Based Sensors

The BJT-based bandgap temperature sensor is the most common type of temperature sensor employed in CMOS technologies. These sensors exploit the fact that ΔV_{BE} (the difference between the base-emitter voltages of two BJTs biased at different current densities) is proportional to the absolute temperature, while the base-emitter voltage (V_{BE}) of a single BJT is complementary to the absolute temperature. This is shown in Figure 1.1 where the current density ratio is n:1. By comparing ΔV_{BE} and V_{BE} to each other ($\Delta V_{BE}/V_{BE}$), the absolute temperature can be determined. These

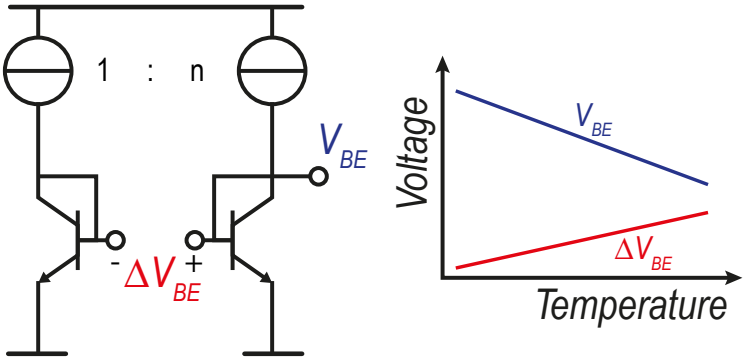

Figure 1.1: Typical BJT-based temperature sensor operating principle.

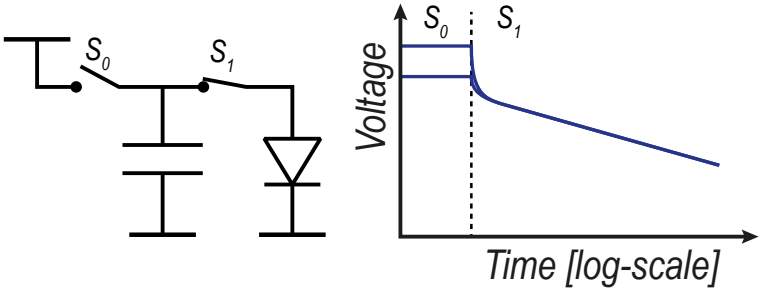

Figure 1.2: Diode/BJT discharing temperature sensing principle.

types of temperature sensors have been around for a long time, and their performance — especially their aging and drift — is quite well understood. In miniaturized versions, a BJT device is sometimes placed remotely to save area (as multiple BJTs for multiple temperature readings share the same readout) [2][4]. However, routing these analog lines is not trivial as they must be well shielded from digital circuit activity.

A downside of BJTs is that they require a large amount of headroom. The required base-emitter voltage can be as high as 0.9 V at low temperatures. On top of that, a current mirror is required (see Figure 1.1), which typically limits BJT-based temperature sensors to above 1.1 V. This is undesirable in the thermal management of SoCs because their operating voltages are typically between 0.6-0.9 V. In [5] an alternative biasing technique is proposed that prevents the extra headroom requirement. It involves discharging a pre-charged cap through a diode (Figure 1.2). As long as the initial voltage exceeds the base-emitter voltage, the discharging voltage

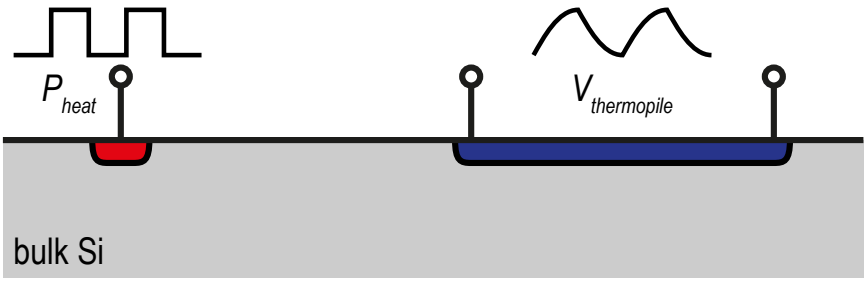

Figure 1.3: Thermal-diffusivity-based temperature sensing principle.

transient is not dependent on the initial charge. The same trick can also be applied using BJTs [6]. This approach does not require current sources on top of the BJTs, so the required headroom may be as low as 0.95 V [6]. In advanced processes, however, this is still higher than the digital supply voltage.

1.2.2 – Thermal Diffusivity-Based Sensors

Thermal diffusivity-based temperature sensors use the propagation delay of heat through the silicon as a measure of temperature [7][8]. Since the silicon that is used in advanced CMOS technologies is very pure and crystalline, its thermal diffusivity (the speed at which heat propagates) is well-defined. Typically, a resistor/switch generates a periodic heat pulse that is picked up by temperature difference detectors (thermopiles) to detect and subsequently digitize the phase shift (Figure 1.3). The advantage is that the accuracy of such sensors is mainly limited by lithography, making them quite accurate, even when untrimmed. The downside is that they require a relatively large amount of energy to generate the heat signal (~3 mW) while not achieving good resolution (~300 mK in 1 ms).

Overall, thermal diffusivity-based temperature sensors are well suited to thermal management applications [1]. They are small, accurate out of the box (little trimming required), can operate from any supply voltage, and scale well with technology. However, their energy efficiency is not yet good enough. When tens of temperature sensors are used on a die, they will consume tens of milli-Watts, which can be problematic for battery-powered devices.

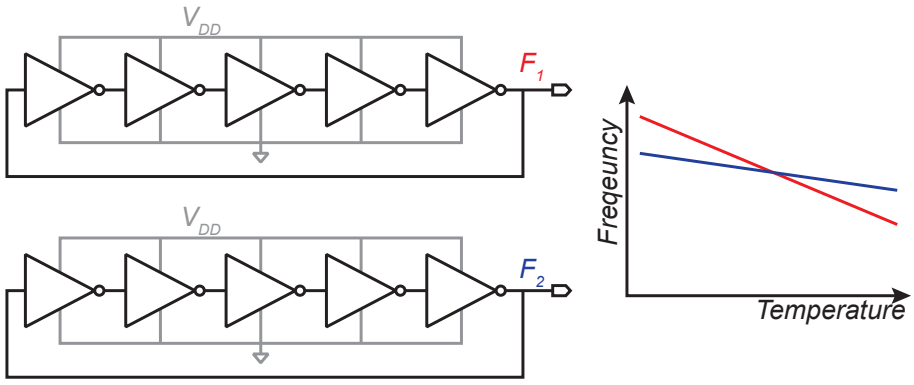

Figure 1.4: Typical MOS leakage-based temperature sensor. The frequency ratio between ROs is used to minimize the supply sensitivity.

1.2.3 – MOS-Based Sensors

MOS-based temperature sensors come in a variety of methods for measuring temperature. One uses the same approach as BJT-based sensors, but with the BJTs replaced by MOS or dynamic threshold MOS (DTMOS) devices [9][10]. The advantage is that these (DT)MOS-based devices do not require the high headroom that BJTs require. A major disadvantage is that they require large area to achieve good stability and accuracy, and they exhibit poor aging compared to BJTs, mainly due to the trapped charges in the oxide/gate-insulator.

Another method is to use the MOS transistor leakage current to measure temperature. This current is exponentially dependent on the temperature and is usually digitized by a ring oscillator (RO) to save area (see Figure 1.4). As a downside, the leakage current is also exponentially dependent on the drain-source voltage (V_{DS}), and also the ring oscillator is highly supply-dependent [11]. Consequently, the main hurdle for these types of temperature sensors is achieving sufficient power-supply rejection to enable operation on the relatively wide and dynamically changing supply range of SoCs. Another bottleneck is that the effect of aging on the temperature accuracy is not well studied but is expected to be much worse than alternatives. This is mainly due to the aging effects of trapped charges in the oxide/gate-insulator that other types of devices, e.g., BJTs, do not experience.

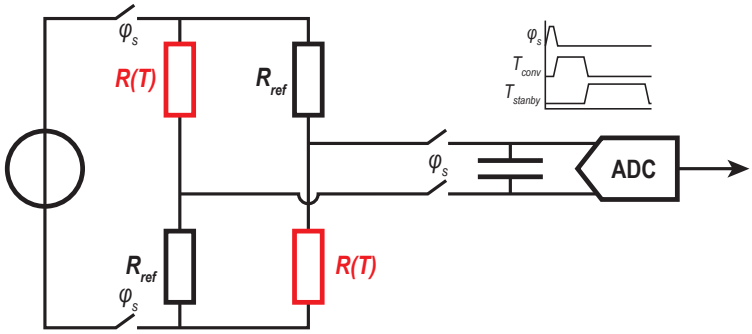

Figure 1.5: Resistor-based temperature sensor [13].

1.2.4 – Resistor-Based Sensors

Resistor-based temperature sensors have become quite popular in recent years [12]. This is mainly because of their high energy efficiency, which is best in class and results from their large temperature coefficient (tempco) and low noise. However, for thermal management, their energy efficiency is already much better than required. On the other hand, accuracy is paramount, which is where resistor-based temperature sensors also do well. They achieve < 1 K accuracy after a 1-pt trim, which is mainly enabled by their trim ability, which is due to their relatively simple underlying physics. On top of that, they are not constrained by voltage headroom and can operate from any voltage.

One example of recent work in resistor-based temperature sensors is [13]. It uses a clocked Wheatstone bridge whose output voltage is sampled and subsequently digitized by a SAR ADC, see Figure 1.5. The Wheatstone bridge employs temperature-dependent resistors and reference (temperature-independent) resistors to generate a temperature-dependent output voltage. During the SAR conversion, the clocked Wien-bridge is disabled to save power.

1.3: Area vs. Accuracy

Figure 1.6 shows the relative inaccuracy (i.e., inaccuracy divided by the temperature range) vs. the area of different types of temperature sensors focusing on designs smaller than 10 000 µm² and Figure 1.7 shows the same but then for trimming points. Both show a trend where smaller

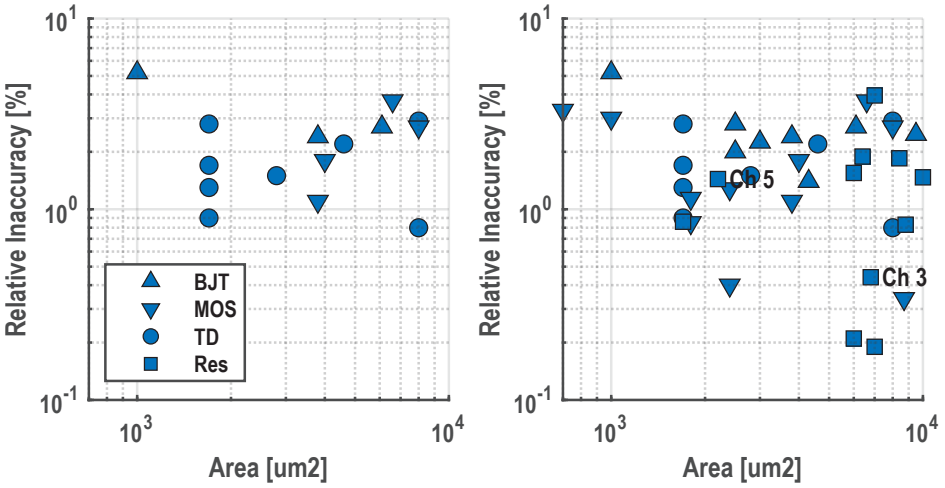

Figure 1.6: Relative inaccuracy vs. area of temperature sensors (a) at the start of this work and (b) after this work (2022-11) [12]

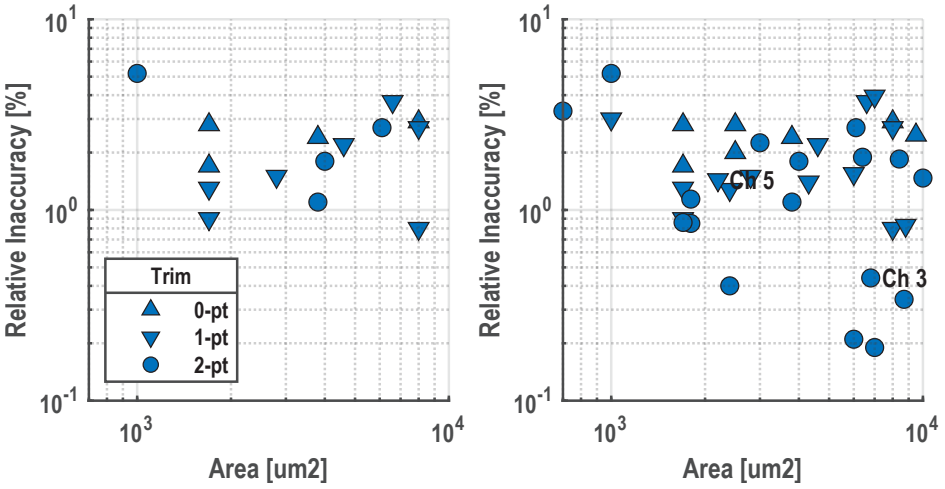

Figure 1.7: Relative inaccuracy vs. area of temperature sensors vs. trim points (a) at the start of this work and (b) after this work (2022-11) [12]

designs tend to be less accurate. Although this does not represent a fundamental tradeoff, there are two main reasons why this happens. The first is that readout circuits in such a small footprint usually contribute towards inaccuracy. To fit in such small footprints, they usually employ time-domain readouts that measure small temperature-dependent time constants from compact frontends. Another reason is that mismatch in minimized sensing elements will start to dominate. Usually, the effect

of mismatch is drowned by the effect of process variation. However, the mismatch eventually becomes significant when the sensing element's area is reduced. The area-vs.-accuracy tradeoff is the main challenge in designing temperature sensors for thermal management. Consequently, designers must be cautious when spending area, especially if there is little accuracy return on the spent area.

1.4: Voltage vs. Time-Domain

While the scaling of CMOS enables smaller dimensions and higher speeds, it comes at the cost of reducing the supply voltage. Consequently, while the voltage headroom is decreasing, the timing resolution of advanced nodes is increasing. This is the main reason why many designs are ported to the time domain [14]. Another benefit of time domain processing is that it tends to require less area [15]. The latter is mainly because voltage domain designs usually require relatively large transistors as well as capacitors. For these reasons, time-domain readouts should be preferred for thermal management targeted temperature sensors. Consequently, this thesis focuses on scalable, highly digital, time-domain readouts.

1.5: Thesis Objective

At the start of the research for this thesis, resistor-based temperature sensors were just gaining popularity because of their high energy efficiency. Although their accuracy and stability were good enough for thermal management, their potential for scaling had not yet been investigated. The goal of this work is to see how well the area of the resistor-based temperature sensors can be scaled without severe loss of accuracy and to see how suitable they are for use in thermal management applications.

1.6: Thesis Contributions

This thesis describes a series of explorations that ultimately resulted in a low area (2,210 μm^2) and low power (28 μW) design that achieves a good resolution of 12.8 mK (rms) in a 1 ms conversion time with a ±1.3°C (3σ) accuracy from -55°C to 125°C. These explorations include the design of compact temperature sensors 1) using CCO-based highly-digital

PDSDM, 2) using the combination of a comparator and a counter-based highly-digital PDSDM, and based on the knowledge acquired from these designs, 3) using a dual GRO + counter-based highly digital PDSDM that resulted in the best performance. In the process, the following issues were investigated: 1) time quantization effects in highly digital PDSDMs (such as dead zones and beat cycles), 2) supply sensitivity of poly-phase filters (especially in combination with static/dynamic comparators), and 3) the dual GRO structure (including the impact of leakage, mismatch, offset, and preferred layout).

1.7: Thesis Organization

Chapter 2 first gives a short overview of the physical properties of the resistors available in CMOS technologies. It then goes on to discuss a few different ways to digitize resistance, eventually settling on a highly digital phase-domain sigma-delta modulator (PDΣΔM). Finally, some of its important non-idealities are discussed, such as time-domain quantization noise and wrap-around.

The following three chapters describe a series of steps culminating in the final design. First, in Chapter 3, the highly digital CCO-based readout from [16] is merged with a modified single-ended version of the Wien-Bridge sensor from [17]. It is shown that the main accuracy limitation of the design arises from the non-linear current-to-frequency transfer of the CCO. Then, in Chapter 4, the CCO is removed from the signal path using a digital mixer to improve the sensor's accuracy. In this case, the counter is clocked at a constant frequency, revealing an underlying problem of the PDΣΔM. This chapter shows that the interaction between the PDΣΔM's counter frequency and the ΣΔM sampling frequency causes timing artifacts that negatively impact the sensor's operation. Finally, in Chapter 5, a dual gated-ring oscillator is employed that maintains zero impact from CCO non-linearity and prevents the impact of these timing artifacts. This is enabled by using two oscillator and counter pairs and improves the energy efficiency of the PDΣΔM by two orders of magnitude.

Finally, Chapter 6 gives a conclusion and future recommendations.

1.8: References

[1] U. Sonmez, R. Quan, F. Sebastiano, and K. A. A. Makinwa, "A 0.008-mm² Area-Optimized Thermal-Diffusivity-Based Temperature Sensor in 160-nm CMOS for SoC Thermal Monitoring," in *ESSCIRC 2014 - 40th European Solid State Circuits Conference (ESSCIRC)*, 2014.

[2] J. Shor, "Compact Thermal Sensors for Dense CPU Thermal Monitoring and Regulation: A Review," *IEEE Sensors Journal*, vol. 21, no. 11, pp. 12774-12788, 2021.

[3] J. Shor and K. Luria, "Miniaturized BJT-based thermal sensor for microprocessors in 32-and 22-nm technologies," *IEEE journal of solid-state circuits*, vol. 48, no. 11, pp. 2860-2867, 2013.

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

[5] M. Eberlein and H. Pretl, "A No-Trim, Scaling-Friendly Thermal Sensor in 16nm FinFET Usig Bulk Diodes as Sensing Elements," *IEEE Solid-State Circuits Letters*, vol. 2, no. 9, pp. 63-66, 2019.

[6] 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," in *IEEE International Solid-State Circuits Conference*, San Fransisco, 2023.

[7] C. P. van Vroonhoven and K. A. A. Makinwa, "Thermal diffusivity sensors for wide-range temperature sensing," in *SENSORS*, Lecce, Italy, 2008.

[8] C. P. van Vroonhoven and K. A. A. Makinwa, "A CMOS Temperature-to-Digital Converter with an Inaccuracy of ±0.5 C (3σ) from -55 to 125°C," in *IEEE International Solid-State Circuits Conference*, San Fransisco, 2008.

[9] K. Souri, Y. Chae, Y. Ponomarev, and K. A. A. Makinwa, "A precision DTMOST-based temperature sensor," in *ESSCIRC*, 2011.

[10] Y. Kim, W. Choi, J. Kim, S. Lee, S. Lee, H. Kim, K. A. A. Makinwa, and Y. Chae, "A 0.02mm² embedded temperature sensor with ±2°C inaccuracy for self-refresh control in 25nm mobile DRAM," in *ESSCIRC Conference*, 2015.

[11] Z. Tang, Y. Fang, Z. Shi, X.-P. Yu, N. N. Tan, and W. Pan, "A 1770-μm² Leakage-Based Digital Temperature Sensor With Supply Sensitivity Suppression in 55-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 3, pp. 781-793, 2020.

[12] K. A. A. Makinwa, "Temperature Sensor Performance Survey," [Online]. Available: http://ei.ewi.tudelft.nl/docs/TSensor_survey.xls. [Accessed 20 November 2022].

[13] H. Xin, M. Andraud, P. Baltus, E. Cantatore, and P. Harpe, "A 174 pW–488.3 nW 1 S/s–100 kS/s All-Dynamic Resistive Temperature Sensor With Speed/Resolution/Resistance Adaptability," *IEEE Solid-State Circuits Letters*, vol. 1, no. 3, pp. 70-73, 2018.

[14] M. Baert and W. Dehaene, "A 5-GS/s 7.2-ENOB Time-Interleaved VCO-Based ADC Achieving 30.5 fJ/cs," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 6, pp. 1577-1587, 2020.

[15] G. Taylor and I. Galton, "A Reconfigurable Mostly-Digital Delta-Sigma ADC With a Worst-Case FOM of 160 dB," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 4, pp. 983-995, 2013.

[16] R. Quan, U. Sonmez, F. Sebastiano, and K. A. A. Makinwa, "A 4600μm² 1.5°C (3σ) 0.9kS/s thermal-diffusivity temperature sensor with VCO-based readout," in *IEEE International Solid-State Circuits Conference*, San Fransisco, 2015.

[17] S. Pan, Y. Luo, S. H. Shalmany, and K. A. A. Makinwa, "A Resistor-Based Temperature Sensor With a 0.13 pJ·K² Resolution FoM," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 1, pp. 164-173, 2018.

Chapter 2 | Resistor-Based Temperature Sensors

There are many types of resistors that are available in CMOS technology and many types of readout circuits that can be used to digitize their resistance. Not all resistors and readouts, however, can be used to realize a temperature sensor with desirable properties. For thermal management, these sensors must be small (<<10 000 µm²), have moderate accuracy (~1 K, -55—125 °C), have no or little cross-sensitivities (resulting in temperature errors smaller than its accuracy), and achieve a resolution of ~100 mK in a bandwidth of 1 ms with negligible power consumption (few mWs). This chapter describes the types of resistors that are available in CMOS and which of them are best suited for thermal management. Then the chapter discusses the different types of readouts that can be used to digitize resistance. In particular, the highly digital phase-domain sigma-delta modulator (PDΣΔM) proves promising for thermal management. Finally, a few advanced aspects of its operation are covered.

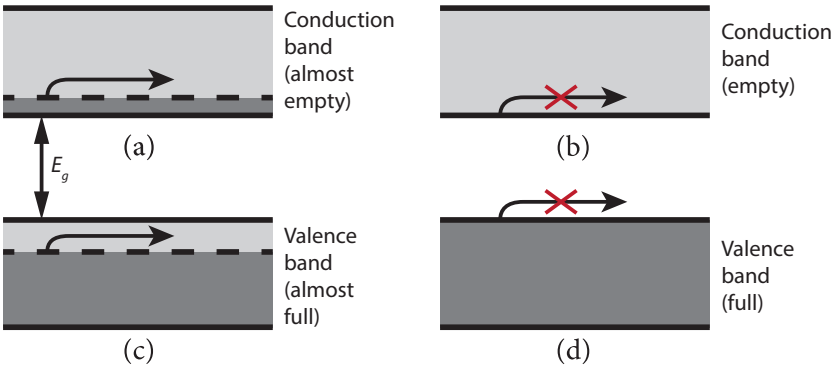

Figure 2.1: Electrical conduction/isolation when (a) free states are available in the conduction band , (b) no charge carriers are available in the conduction band, (c) free states are available in the valence band, and (d) no free states are available in the valence band.

2.1: Electrical Conduction

To better understand the merits and behavior of different types of resistors, first a short summary of electrical conduction is given. This is not meant as a comprehensive study, more in-depth analysis of this topic can be found in [1–2].

Electrons experience a mechanical force when subjected to an electric field, which can cause the electrons to change their state and accelerate. However, electrical conduction in solid-state materials can only occur when there are free states available for the electrons to move into, i.e., when they are in the conduction band (Figure 2.1 (a)) or when there are free states in the valence band (*holes*)(Figure 2.1 (c)).

Since metals have either Fermi levels within the conduction band or overlapping conduction and valence bands, there are always free states available to allow conduction. Undoped semiconductors, on the other hand, have Fermi levels in the bandgap; thus, no conduction is possible at 0 K. At higher temperatures, the Fermi-Dirac function widens, which allows small numbers of charge carries to conduct electricity. Doping is added to move the Fermi level closer to either the conduction or valence band, which drastically increases the charge carrier concentration and, therefore, the electrical conduction. The total current density is modeled by $J = v \cdot e \cdot n$, where v is the charge velocity, and n is the charge carrier concentration.

Under the influence of an electric field, electrons will accelerate and collide with obstacles along their path, which limits their velocity. The electric field strength and the mean free path between collisions determine the effective electron velocity. However, this collision process is highly statistical and complex. Consequently, it is modeled by the empirical model: $v = \mu E = e\tau E/m$, where τ is the mean time between collisions, μ is the electron/hole mobility, and m is the effective mass [1].

There are two collision mechanisms that dominate in semiconductor materials: 1) *phonon scattering* and 2) *ionized impurity scattering*. The former is due to vibrations in the crystal lattice, which happens at temperatures above 0 K. Higher temperatures cause more vibrations, which generally result in a higher probability of electron-phonon scattering. To first order $\mu_L \propto T^{-3/2}$, but in practice $\mu_L \propto T^{-n}$, where n is 2.3 for electrons and 2.1 for holes [3]. The latter is due to collisions with (charged) impurities in the lattice. These can arise from either unintended imperfections or the doping of the silicon. In theory, for low to medium impurity concentrations, *ionized impurity scattering* starts dominating at temperatures below 100 K [4]. In general, impure silicon has lower conductivity and its conductivity is less temperature-dependent.

Another collision mechanism observed in materials is *surface scattering*, which happens when charge carriers scatter against the edges of the conductor. This type of scattering only becomes significant as the thickness of the conductor approaches the mean free path of the charge carriers. For silicon, this distance is ~1.95 µm, so silicon conductors with a thickness close to 2 µm or less show a noticeable decrease in electron mobility.

As a result of the many collision mechanisms that determine a resistor's temperature dependence, this is usually quite non-linear. These mechanisms also result in significant resistance spread, which then requires multi-point calibration to mitigate their impact. Consquently, resistor-based temperature sensors often require polynomial correction to remove systematic non-linearity, and multi-point calibration to remove part-to-part inaccuracies.

Table 2.1: Resistor types and their typical properties.

resistor type	collision mechanism	TC¹ [mΩ/Ω/K]		TC² [µΩ/Ω/K²]		Rsq [Ω/sq]	
		min	max	min	max	min	max
Interconnect	*phonon, surface scattering*	2.47	3.29	-1.62	-0.25	0.02	0.25
Diffusion	*phonon, impurity, surface scattering*	1.33	1.57	0.76	0.95	97.99	707.40
Poly-Silicon	*impurity, phonon, surface scattering*	-0.36	-0.02	0.36	0.72	129.23	817.31
Silicide	*phonon, surface scattering*	2.08	2.34	-1.24	-0.69	8.77	18.39

In the next sub-sections, the different types of available resistors in CMOS technologies will be discussed. Table 2.1 shows an overview of the available resistor types and their typical properties. Figure 2.2 (next page) shows a cross-section of the different resistor types.

2.1.1 – Interconnects

Interconnects are used to connect on-chip components to each other and with the outside world. Ideally, since interconnects should not contribute to the electrical behavior of the circuit, their electrical resistance should be as small as possible. Consequently, interconnects are usually made of aluminum and copper, which are good conductors. Electrons in copper have a mean free path that is approximately 40 nm [5]. While this is irrelevant for mature processes which have a minimal metal width that is much wider than 40 nm, due to aggressive scaling, advanced nodes are affected by *surface scattering*.

Due to their many requirements, interconnects make lousy thermistors for thermal management. Their high conductivity means that a very long (and narrow) geometry is required to achieve resistances within a measurable range (larger than a few tens of kΩ). This requires a significant amount of area, especially in more mature technology nodes where the minimum interconnect width is relatively wide. However, the low electron mean free path of copper and its high charge carrier availability result in a conductor whose resistance is purely defined by *phonon scattering*. Therefore, interconnect resistors are expected to be some of the most accurate resistors available in CMOS.

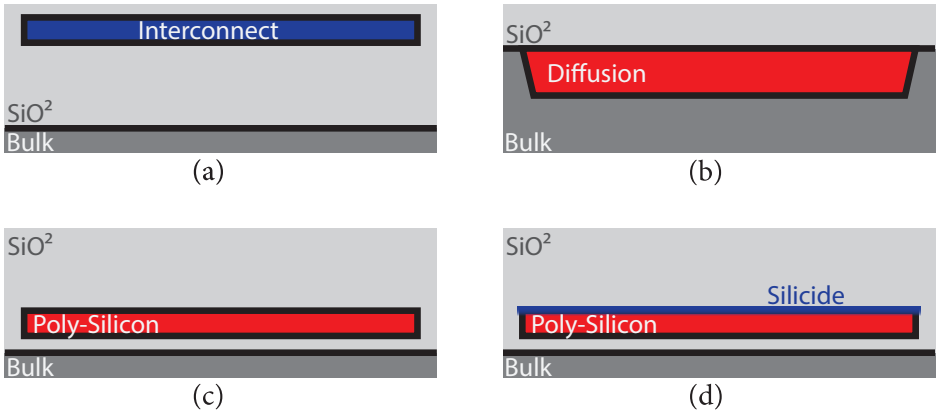

Figure 2.2: Cross section of different resistor types; (a) interconnect, (b) diffusion, (c) poly-silicon, and (d) silicide.

Recently, a temperature sensor (for thermal management) using an interconnect resistor as its sensing element was proposed [6]. To achieve a resistance of 121.7 kΩ, a width/length of 0.02 µm/1500 µm is used. These dimensions are enabled by a 5 nm technology. Interestingly, the accuracy of the proposed sensor (4 K from -55 to 125 °C, 1-pt trim, 3σ) is not at par with the state-of-the-art. One reason could be that the interconnect width (20 nm) is smaller than the expected electron mean free path in copper (40 nm). However, it is much more likely that the accuracy is limited by the chosen readout circuit.

2.1.2 – Diffusion

The temperature coefficient (TC) of diffusion resistors depends on both *impurity scattering* and *phonon scattering*. Diffusion resistors require doping to create the charge carriers needed for electrical conduction. In turn, these dopant atoms cause impurities in the crystal lattice, resulting in reduced conduction, especially at lower temperatures. As a result, their (higher-order) TC depends, among other factors, on the dopant concentration. This makes diffusion resistors more difficult to trim, as more temperature points are required for accurate trimming. On top of that, diffusion resistors have a charge depletion region at their interface with the p-sub/n-well, which causes leakage currents, voltage sensitivity, and noise coupling from the substrate. As a result, diffusion resistors are undesirable for thermal management applications.

2.1.3 – Poly-Silicon

Poly-silicon is an often used gate material that is grown on top of the gate insulator. It is composed of crystalline silicon grains with boundaries between the grains. These boundaries contribute to the resistance and have a negative TC. Furthermore, these boundaries contribute to a voltage coefficient (VC) [7] which can impact supply sensitivity, thus degrading cross-sensitivity and ultimately accuracy. The resulting TC of poly-silicon is determined by a combination of a positive (due to *phonon scattering*) and negative TC (due to *impurity/surface scattering*), resulting in a low TC that depends on multiple parameters. They are also known to be affected by aging (due to electromigration)[8]. Consequently, poly-silicon resistors are undesirable for use in temperature sensors.

2.1.4 – Silicide

To contact a diffusion resistor, an ohmic contact between metal & silicon must be made. To prevent the formation of a Schottky diode, which appears at semiconductor/metal interfaces, a layer of silicide is annealed into the silicon lattice, resulting in a compound material. This silicide layer is used on all semiconductor/metal interfaces, i.e., on top of diffusion and on top of poly. Except for the n-well resistor, all resistors are available with this silicide layer grown over the entire surface. In fact, if unsilicided resistors are required, a special silicide stop layer is necessary to locally prevent the growth of silicide.

The resistance of most silicides exhibit near-linear metal-like behavior over temperature. They only deviate from this linear behavior at extremely low temperatures, i.e., below the Debye temperature (typically < 200 K). Some silicides, just like some metals, can also show resistance saturation at higher temperatures (typically > 500 K) [9].

Since silicides exhibit metal-like behavior, metal-like thermistor performance can be achieved without the need for long and narrow interconnects. The main reason for this is that the silicide layer is very thin,

so the sheet resistance of silicided resistors is much higher than that of the metal interconnects. Of the two types of silicided resistors, silicided-poly resistors are the most attractive, as they do not contain a charge depletion region interface with the substrate. Although their behavior is mostly metallic, they have a small VC due to the grain boundaries of the silicide. Also, their effective electron mean free path seems to be around 0.2um [10]. Below this width, their sheet resistance will heavily depend on its width [11]. In principle, they should also be susceptible to aging (through diffusion and electromigration). However, an aging experiment (Chapter 5) indicates that after an initial 1-point trim, their aging does not significantly impact sensor accuracy. Another experiment [12] shows that after 500 hours of rapid aging (at 125°C), the error is smaller than 0.2 K, matching the results of Chapter 5. However, rapid aging measurements of over 10 000 hours using multiple parts are required to gain higher confidence level in silicided poly resistor aging.

Recently, resistor-based CMOS temperature sensors have garnered interest. This is because the high TC of the near-metallic silicided-poly resistor allows it to achieve superior energy efficiency. At the same time, the near metallic silicided-poly resistor has been shown to achieve reasonable accuracy after a 1 or 2-pt trim. This is because its TC is mostly defined by phonon-scattering, resulting in low spread, which allows easy trimming.

2.2: Readout of Resistors

There are two main approaches to reading out resistors: 1) in the amplitude domain and 2) in the time domain. Both cases require a reference to which the resistance is compared. The type of reference that is required and the way digitization takes place differ between these two approaches and have a large impact on the performance.

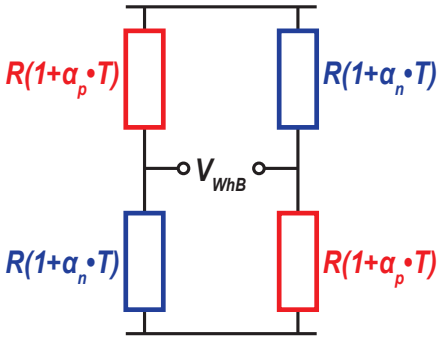

Figure 2.3: Wheatstone bridge temperature sensor.

2.2.1 – Amplitude Domain

The amplitude domain is the most straightforward domain for reading out a resistor. It involves applying a current to the resistor and then digitizing the voltage across it. This approach, which is used by most digital multimeters (DMMs), requires both an accurate current source as well as an accurate voltage source. Since those references require temperature compensation, this approach is not well suited for compact on-chip solutions.

Alternatively, a reference resistor can be used to compare the temperature-sensitive resistor (thermistor) to. The most popular method is to use a Wheatstone bridge configuration (see Figure 2.3), which can achieve superior energy efficiency, as low as 10 fJK² [13]. In fact, this is quite close to the theoretical limit, which is derived from the bridge noise:

$$v_n{}^2 = \frac{4kTR}{2T_i} \tag{2.1},$$

where k is the Boltzmann constant, T is the temperature, R is the effective Wheatstone-bridge resistance, and T_i is the integration time. Since the bridge output V_{bridge} can be expressed as:

$$V_{bridge} = \frac{R_p - R_n}{R_p + R_n} V_{dd} = \frac{\alpha_p T - \alpha_n T}{2 + \alpha_p T + \alpha_n T} V_{dd} \tag{2.2},$$

the bridge sensitivity is given by:

$$\frac{\partial V_{bridge}}{\partial T} = \frac{2(\alpha_p - \alpha_n)}{\left(2 + T(\alpha_p + \alpha_n)\right)^2} V_{dd} \approx 0.5(\alpha_p - \alpha_n) V_{dd} \tag{2.3},$$

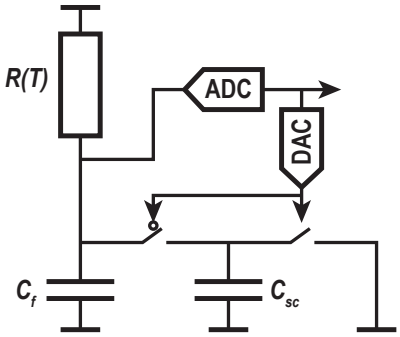

Figure 2.4: Temperature-to-digital converter using switched-cap resistor DAC.

where α_p and α_n are the first-order TCs. The power consumption is equal to ${V_{dd}}^2/R$. Neglecting the power consumption of the readout circuit and its noise contribution, the theoretical energy efficiency is:

$$\frac{8kT}{(\alpha_p-\alpha_n)^2} \quad [\mathrm{JK^2}] \ (2.4).$$

Considering a TC of 0.3 %/K, which is that of a silicided-poly resistor, we reach an ideal energy efficiency of 3.8 fJK².

The main difficulty with realizing a Wheatstone bridge sensor is that there are no accurate untrimmed or temperature-stable resistors available in CMOS technology. To maximize temperature sensitivity, Wheatstone bridge sensors usually employ opposite TC resistors. As a result, their noise resolution is improved compared to their temperature-stable reference counterpart. However, as a downside, both resistors contribute towards TC spread and inaccuracy. Therefore, these types of sensors are typically less accurate (0.1—0.5 K) than their single TC counterparts (0.03—0.11 K). As discussed previously, the main difficulty in thermal management is to design sensors that are both accurate as well as small. Therefore, it is undesirable to spend twice the area (two resistors) while also doubling the inaccuracy.

As a halfway house between the amplitude and time domain approaches, a switched capacitor resistor can be used as a temperature-stable resistor [14] (see Figure 2.4). This solution does not suffer from the increased inaccuracy of the dual resistor approach. Furthermore, its output is still a DC output, allowing for good energy efficiency. However, to maintain low

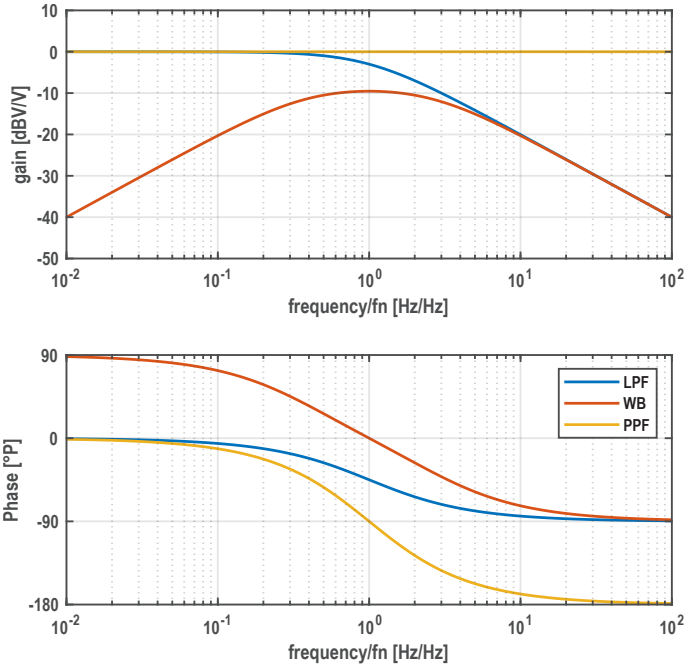

Figure 2.5: LPF, WB, and PPF filter responses; PPF around the nominal frequency has the steepest phase slope.

ripple, a large filter cap (C_f) is required, which occupies a large amount of area. This comes on top of the area that the analog readout already requires to digitize the temperature. Consequently, this approach is not suitable for use in compact temperature sensors.

2.2.2 – Time Domain

While voltage headroom is reducing in the advanced process nodes where thermal management is required, their speed is increasing as well. Therefore, a better approach for thermal management is the use of a time-domain based readout circuit. This means that scaling/porting time-domain readouts to smaller technologies automatically results in improved timing resolution, accuracy, and energy efficiency.

In a time-domain readout circuit, a resistor is combined with a dynamic element to form a time constant. A capacitor is usually used as the dynamic element since it can be dense and temperature-stable. Instead of a current/voltage reference or resistor reference, this system uses a time reference

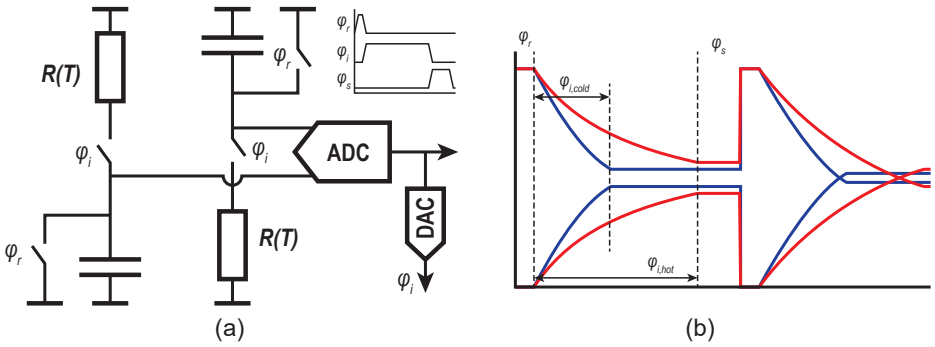

Figure 2.6: (a) simplified block diagram of an-LPF based temperature sensor [13] and (b) the transient input of the ADC, which tunes φ_i so its input is zero during φ_s.

to digitize the RC time constant. As a benefit, accurate and temperature-stable time references are readily available in the dense digital circuitry that requires thermal management.

There are three main topologies that comprise capacitors and resistors to make a temperature-dependent time constant. They are the first-order low-pass filter (LPF), the Wien-bridge (WB), and the poly-phase filter (PPF). The phase responses of these filters to a sine-wave input signal are plotted in Figure 2.5. Even though, for complexity reasons, these filters are usually driven by square waves (especially in the case of thermal management), their phase response plots are still a good predictor for phase sensitivity. The next three sub-sections will explore the pros and cons of each of these filter topologies.

2.2.2.1 – Low-Pass Filter

The LPF is the most basic filter topology of the three. Although it requires fewer components than the WB (1 capacitor and 1 resistor), its sensitivity to resistance, and therefore temperature, is the lowest. Consequently, the WB is the preferred filter when directly digitizing the phase response because of its improved temperature sensitivity, reduced power-supply sensitivity, and its band-pass filter.

One published design, which is based on an LPF topology, samples the LPF voltage during its transition to determine the time constant [15] (see Figure 2.6). This is only possible with an LPF since it is composed of only

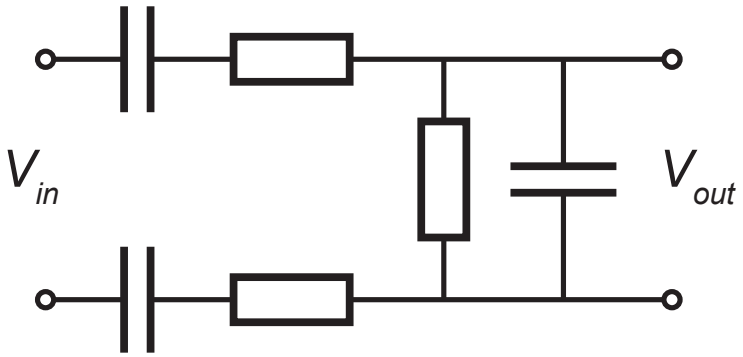

Figure 2.7: Wien bridge Filter.

a single capacitor that can easily be sampled. The sampled noise is equal to kT/C, where k is the Boltzmann constant, T is the temperature, and C is the capacitor value. Its sensitivity is:

$$\frac{\partial V_s}{\partial T} = \frac{V_{dd} \cdot T_d \cdot \alpha \cdot e^{-T_d / R_0(1+\alpha T_d)C}}{R_0(1+\alpha T)^2 C} \approx \frac{V_{dd} \cdot T_d \cdot \alpha \cdot e^{-T_d / R_0 C}}{R_0 C} \quad (2.5),$$

where V_s is the sampled voltage, V_{dd} is the supply voltage, T_d is the sampling moment, which is around the halfway crossing point (i.e., $T_d \approx \ln(2) R_0 C$), α is the temperature coefficient of the resistor, and R_0 is the resistance at room temperature. This results in:

$$\frac{\partial V_s}{\partial T} \approx \frac{V_{dd} \cdot \ln(2) \cdot \alpha}{2} \quad (2.6).$$

The consumed energy is $0.5 \cdot {V_{dd}}^2 \cdot C$ and the resulting energy efficiency is:

$$FoM \approx \frac{2kT}{\ln(2)^2 \cdot \alpha^2} \quad (2.7).$$

For α is 0.3 %/K, the ideal FoM is 1.9 fJK². Surprisingly, this is better than the theoretical FoM of the Wheatstone bridge. However, only the noise caused by the sampling of the capacitors is taken into account. The power of the readout circuit, which is now a discrete time converter, is not.

2.2.2.2 – Wien Bridge

The WB combines a first-order LPF and a first-order high-pass filter (Figure 2.7). Compared to the LPF, it requires double the components, but also has double the phase sensitivity (Figure 2.5). One benefit of its inherent LPF is that the DC output voltage is insensitive to supply voltage variations. Therefore, the WB can be used in a single-ended setup (see Chapter 3), reducing the area at little cost to accuracy. The single-ended WB requires the same number of components as the differential first-order LPF, but it has double the sensitivity, resulting in improved accuracy.

The WB is typically readout by a PDΣΔM (Ch. 2.3.2), which has decreased energy efficiency compared to the amplitude domain resistor readout. Assuming a sinewave input, the amount of charge that the 1st stage integrator integrates during each ΣΔM cycle can be expressed as:

$$\begin{aligned}\Delta Q &= \int_0^{T_{drive}/2} [A_I \sin(2\pi f_{drive} t + \varphi_{WB}) + I_{os}]dt \\ &\quad - \int_{T_{drive}/2}^{T_{drive}} [A_I \sin(2\pi f_{drive} t + \varphi_{WB}) + I_{os}]dt \\ &= \frac{2A_I \cos(\varphi_{WB})}{\pi f_{drive}} \end{aligned} \quad (2.8),$$

where T_{drive} is the period of the signal driving the WB, which is equal to the ΣΔM period, A_I is the amplitude of the current flowing out of the WB, f_{drive} is the frequency of the signal driving the WB, t is the time, φ_{WB} is the WB phase compared to the PDΣΔM phase, and I_{os} is the offset current that is integrated in the PDΣΔM. The integrated noise in the same period is:

$$\Delta Q_n = T_i \sqrt{\frac{4kT}{R}\frac{6}{9}\cdot\frac{1}{2T_i}}\Bigg|_{T_i = 1/f_{drive}} = \sqrt{\frac{4kT}{R}\frac{6}{9}\cdot\frac{1}{2f_{drive}}} \quad (2.9),$$

where T_i is the integration time, which is equal to $1/f_{drive}$. The WB phase sensitivity at its corner frequency is $2/3 \cdot \alpha$, where α is the temperature coefficient of the resistor. Its temperature noise is then:

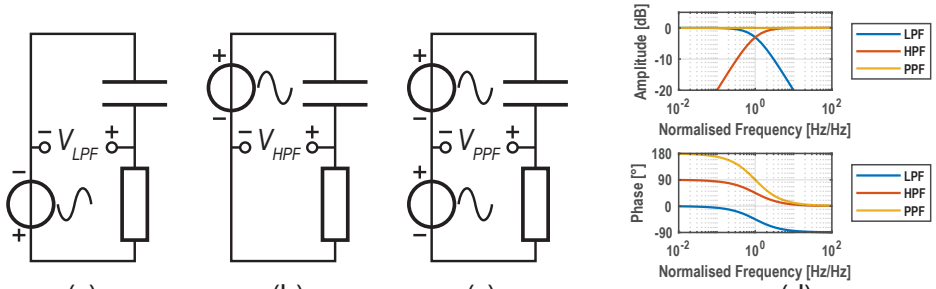

Figure 2.8: Low-pass filter (a), high-pass filter (b), poly-phase filter (c), and their amplitude and phase response (d).

$$T_{n,WB} = \frac{\partial T}{\partial \varphi_{WB}} \cdot \frac{\partial \varphi_{WB}}{\partial Q} \cdot \Delta Q_n = \frac{3}{2\alpha} \cdot \frac{\pi f_{drive}}{2A_I} \sqrt{\frac{4kT}{R}\frac{6}{9} \cdot \frac{1}{2f_{drive}}}$$

$$= \frac{3\pi}{4\alpha A_I} \sqrt{\frac{4kT}{R}\frac{6}{9} \cdot \frac{f_{drive}}{2}} \qquad (2.10).$$

Its RMS power consumption is $A_I{}^2 \cdot R \cdot 3\sqrt{2}/\sqrt{2}$, and so its FoM is:

$$\frac{A_I{}^2 \cdot R \cdot 3}{f_{drive}} \cdot \frac{9 \cdot \pi^2}{16 \cdot \alpha^2 A_I{}^2} \cdot \frac{4kT}{R}\frac{6}{9} \cdot \frac{f_{drive}}{2} = \frac{9 \cdot \pi^2 \cdot kT}{4 \cdot \alpha^2} \qquad (2.11).$$

For a TC of 0.3 %/K, this results in an FoM of 5.9 fJK². However, this assumes a sine-wave input, while usually a square wave is used. Moreover, it neglects the readout, which requires enough bandwidth to minimize its phase-shift contribution. As a result, the effective energy efficiency of WB types of temperature sensors (110 fJK² [16]) is lower than Wheatstone bridge types of temperature sensors (10 fJK² [13]).

2.2.2.3 – Poly-Phase Filter

As shown in Figure 2.8, a PPF requires only one resistor and one capacitor to achieve a second-order phase response. This is achieved by driving the RC filter with anti-phase signals. When the resistor side is driven, the resulting response is that of an LPF (Figure 2.8 (a)). However, when the

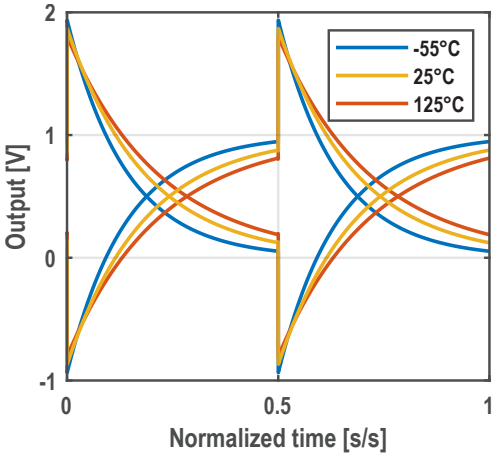

Figure 2.9: PPF transient response for different temperatures and VDD = 1V.

capacitor side is driven, the result is that of a high-pass filter (Figure 2.8 (b)). A superposition of these two filters driven in anti-phase (Figure 2.8 (c)) then results in the phase response of a second-order all-pass filter (Figure 2.8 (d)).

Figure 2.9 shows the PPF output when driven by a square wave for different temperatures. It can be seen that the phases of its zero-crossings (ZCs) are strongly temperature-dependent, making it ideal for readout by a comparator. However, it can also be seen that its output swings beyond the supply rails, which makes the design of the comparator even more challenging.

The use of ZC detection exacerbates the filter's output noise compared to readout circuits in which the filter's output is directly integrated [17]. In a continuous time PDΣΔM, the noise bandwidth, which is determined by the integration time, i.e., $1/(2 \cdot T_{conv})$, can be very narrow. In contrast, the comparator directly translates the instantaneous voltage noise of the PPF's

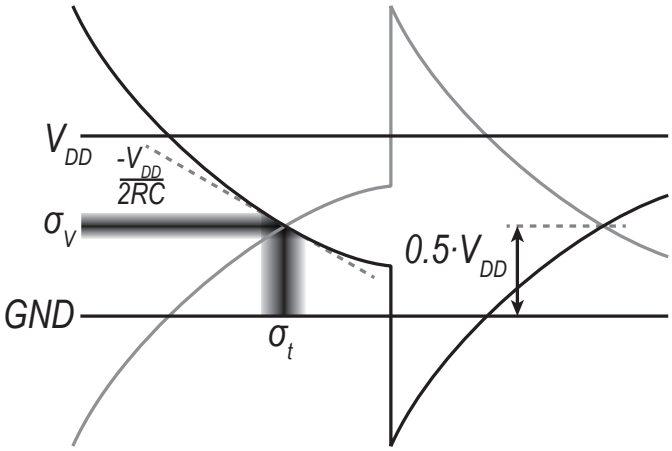

Figure 2.10: PPF's noise sensitivity at zero-crossing.

output into a phase error that depends on the slope of the output signal at the ZCs, as shown in Figure 2.10. The slope, and in turn the sensitivity, of the single-ended PPF at the ZC is given by:

$$\left.\frac{\partial \cdot 0.5 \cdot V_{DD} \cdot e^{-t/\tau}}{\partial t}\right|_{t=0} = \frac{-V_{DD}}{2 \cdot \tau} \tag{2.12},$$

where V_{dd} is the supply voltage, t is the time, and τ is the PPF's RC time constant. The voltage noise of the PPF's output (σ_V) is related to kT/C, while the differential PPF has a $\sqrt{2}$x noise improvement over the single-ended version:

$$\sigma_t = \frac{2 \cdot \tau}{V_{DD}} \cdot \sqrt{\frac{1}{2} \cdot \frac{kT}{C}} \tag{2.13},$$

where C is the PPF capacitor value, k is the Boltzmann constant, and T is the temperature. The phase noise related to each comparator edge is then found to be:

$$\sigma_P = \frac{360°}{T_s} \cdot \sigma_t \tag{2.14},$$

where T_s is the modulator period. There are two edges during each period (T_s), while the total number of periods depends on the total conversion time (T_{conv}). Therefore, the total in-band phase noise for a given conversion time can be found with:

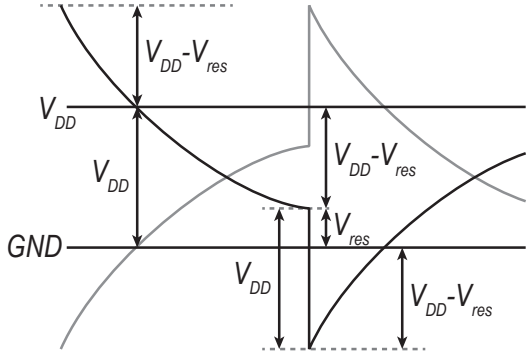

Figure 2.11: Steady-state PPF output waveform.

$$\sigma_{P,IB} = \sqrt{\frac{T_S \cdot 2}{T_{conv}}} \cdot \sigma_P = \frac{2 \cdot R \cdot C}{V_{DD}} \cdot \frac{360°}{T_S} \cdot \sqrt{\frac{kT}{2C} \cdot \frac{T_S \cdot 2}{T_{conv}}} = \frac{720 \cdot R}{V_{DD}} \cdot \sqrt{\frac{kT \cdot C}{T_{conv} \cdot T_S}} \quad (2.15),$$

where R is the PPF resistor value.

The ZC phase, along with the PPF amplitude and power, can be calculated with the help of Figure 2.11. At the end of each half period, $V_{res} \cdot C$ charge is left in the capacitor just before the driving voltages toggle. This means that the maximum amplitude is $2 \cdot V_{DD} - V_{res}$. At steady state, both the initial and final voltages of each half-period are related:

$$V_{PPF}(T_s/2) = V_{res} = (2V_{DD} - V_{res}) e^{-T_s/2\tau} \quad (2.16),$$

which subsequently results in:

$$V_{res} = \frac{e^{-T_s/2\tau}}{1 + e^{-T_s/2\tau}} \cdot 2 \cdot V_{DD} \quad (2.17).$$

The ZC moment happens when the PPF output is equal to half the supply voltage, therefore:

$$V_{PPF}(T_{ZC}) = 0.5 \cdot V_{DD} = (2 \cdot V_{DD} - V_{res}) \cdot e^{-T_{ZC}/\tau} \quad (2.18),$$

which results in:

$$T_{ZC} = \tau \cdot \ln \frac{4}{1 + e^{-T_s/2\tau}} = R \cdot C \cdot \ln \frac{4}{1 + e^{-T_s/2 \cdot R \cdot C}} \quad (2.19).$$

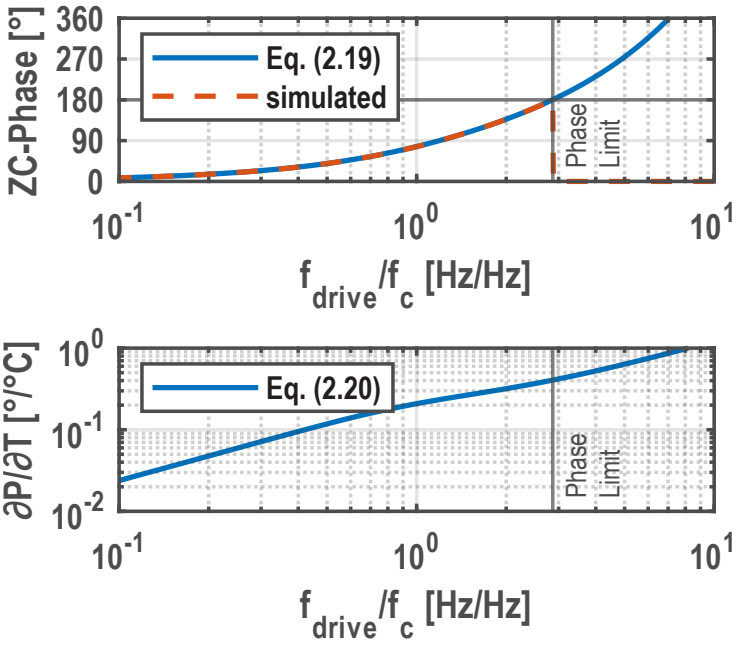

Figure 2.12: ZC phase over normalized driving frequency (top) and PPF phase sensitivity over normalized driving frequency (bottom).

Using Eq. (2.14), the ZC delay is translated into a phase, which is plotted in Figure 2.12 (top). Then, using Eq. (2.19) and under the assumption that the resistor dominates the temperature sensitivity, we can find the phase sensitivity:

$$\frac{\partial P}{\partial T} = \frac{\partial}{\partial R(T)} \cdot \frac{360}{T_s} \cdot R(T) \cdot C \cdot \ln \frac{4}{1 + e^{-T_s/2 \cdot R(T) \cdot C}} \cdot \frac{\partial R(T)}{\partial T}$$

$$= \frac{360 \cdot \tau \cdot \alpha}{T_s} \cdot \left(\ln \left[\frac{4}{1 + e^{-T_s/2\tau}} \right] - \frac{T_s \cdot e^{-T_s/2\tau}}{2 \cdot \tau \cdot \left(1 + e^{-T_s/2\tau}\right)} \right) \qquad (2.20),$$

which is shown in Figure 2.12 (bottom), for $\alpha = 0.3\,\%/°C$, which is the temperature coefficient of silicided p+ poly-silicon. This shows that to maximize the sensitivity, the driving frequency should also be maximized. However, a phase delay of larger than 180° means there are no ZCs before the driving signal (Fdrive) toggles. The consequence of this can be seen in Figure 2.12 (top, simulated), where the measured phase is equal to the driving signal for ZC delays of larger than 180°. Therefore, the driving frequency should remain below 2.8 times the PPF corner frequency.

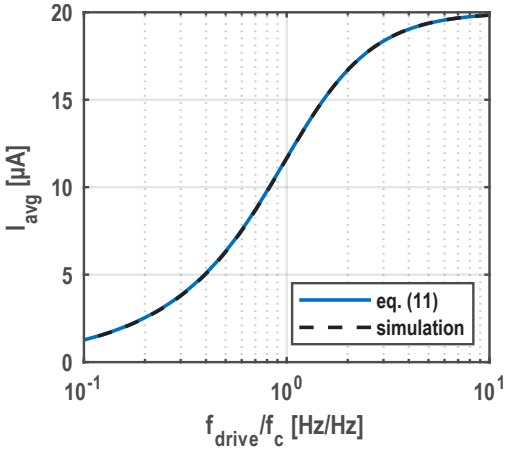

Figure 2.13: Average PPF supply current over normalized driving frequency.

The power consumption of the PPF can be calculated by considering the charging and discharging of its capacitors. Every half period, the capacitors are (dis)charged from $2 \cdot V_{DD} - V_{res}$ to V_{res}, which results in a total transferred charge every half-period of:

$$Q_{hp} = 2 \cdot (V_{DD} - V_{res}) \cdot C \tag{2.21}.$$

This then results in an average current of:

$$I_{avg} = 2 \cdot {Q_{hp}}/{0.5 \cdot T_s} = \frac{8 \cdot V_{DD} \cdot C}{T_s} \cdot \frac{1 - e^{-T_s/2\tau}}{1 + e^{-T_s/2\tau}} \tag{2.22},$$

for a differential PPF, see Figure 2.13.

Finally, the ideal resolution FoM is calculated using (2.20), (2.15), and (2.22):

$$FoM = \left(\frac{\partial T}{\partial P} \cdot \sigma_{P,IB}\right)^2 \cdot I_{avg} \cdot V_{DD} \cdot T_{conv} = \frac{2^2 \cdot 8 \cdot kT}{\alpha^2} \cdot \frac{1 - e^{-T_s/2\tau}}{1 + e^{-T_s/2\tau}} \cdot \left(\ln\left[\frac{4}{1 + e^{-T_s/2\tau}}\right] - \frac{T_s}{2 \cdot \tau} \cdot \frac{e^{-T_s/2\tau}}{1 + e^{-T_s/2\tau}}\right)^{-2} \tag{2.23}.$$

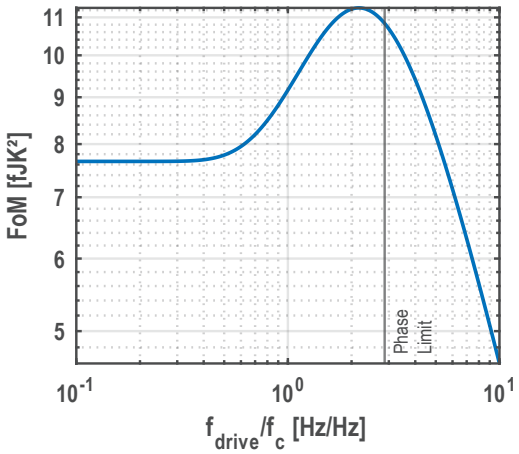

Figure 2.14: Ideal PPF FoM over normalized driving frequency (T = 300 K, α = 0.3 %/°C).

Unsurprisingly, this is independent of the supply voltage or conversion time. It only depends on the temperature, the resistor's temperature coefficient, and the normalized driving frequency, i.e., $F_{drive} = 2\pi\tau/T_s$. Figure 2.14 shows the FoM for T = 300 K and α = 0.3 %/°C as a function of the driving frequency. It can be seen that a FoM as low as 7.7 fJK² can be obtained. This is the ideal, best-case FoM. In reality, driving the PPF and digitizing its phase not only costs energy but also adds noise.

2.3: Miniaturized Temperature-to-Digital Converters

The readout circuit that is used to digitize the thermistor plays a significant part in the performance of the sensor. For thermal management, the readout should be small, use few analog components, and be robust to digital supply noise, simple to design and use, and easy to port to advanced processing nodes. Two types of readout are popular: using a frequency locked loop (FLL) or using a PDΣΔM. The former locks a frequency to the filter's time constant and is essentially a temperature-to-frequency converter, in which case a frequency-to-digital converter is used to digitize the frequency. The latter drives the filter at a constant frequency and directly digitizes its phase shift. The next two sections will take a closer look at each of these.

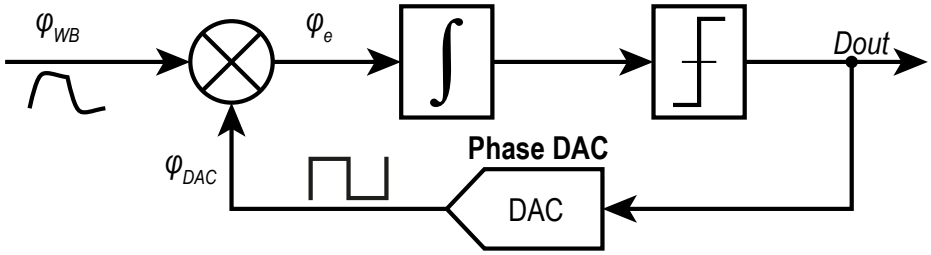

Figure 2.15: Simplified block diagram of a PDΣΔM.

2.3.1 – Frequency Locked Loop

There are two main approaches for making a temperature-to-frequency converter. The first employs a relaxation oscillator to directly generate a frequency that is proportional to the filter's time constant, and hence to temperature [18]. The second approach uses a frequency locked loop (FLL) to lock the frequency of a controlled oscillator to the filter's time constant [19] [20] [15]. In both cases, a consecutive step is required in which the output frequency is digitized. This last step is often neglected in state-of-the-art designs. Assuming that a simple counter can be used to digitize the frequency, the cost of digitization can be quite low. However, to drive the controlled oscillator and make the controlled loop, an analog integrator is required. This integrator involves a capacitor integrator which can occupy a significant proportion of the area. An approach that is more digital and that does not rely on charge (capacitor) based integrators is a better option.

2.3.2 – Phase-Domain Sigma-Delta Modulator

Figure 2.15 shows a simplified block diagram of a PDΣΔM [17]. An input phase (φ_{WB}, at a frequency of F_{WB}) is first multiplied by a feedback phase (φ_{DAC}, also at a frequency of F_{WB}). For sinusoidal signals, this results in a signal φ_e whose DC component is proportional to $\cos(\varphi_{WB} - \varphi_{DAC})$. For phase differences close to 90°, $\cos(\varphi_{WB} - \varphi_{DAC} - 90°) \sim \varphi_{WB} - \varphi_{DAC}$ and so the demodulator effectively performs a phase-domain subtraction. In practice, the higher harmonics present in the WB output and the square wave of the phase DAC output also contribute to φ_e but the deviation is less than 5°. The DC component of φ_e and its harmonics are integrated and then digitized by a 1-bit quantizer whose output is fed back to the phase DAC. This drives the multiplier's DC output to zero so that the average feedback phase φ_{DAC} is equal to φ_{WB} + 90°.

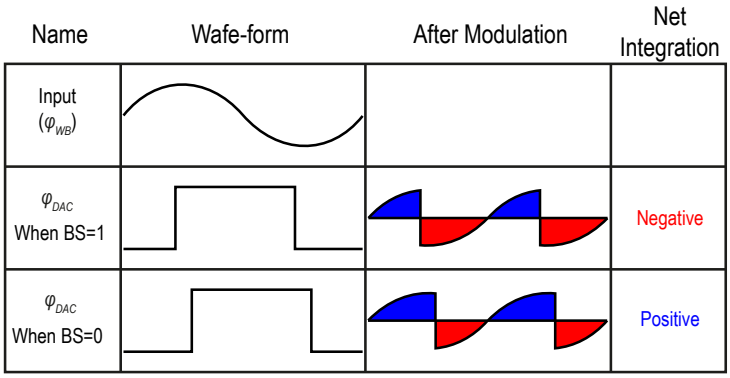

Figure 2.16: Illustration of phase-DAC operation.

Figure 2.16 shows the phase error in the case of a sinusoidal input signal. Here, the phase-DAC generates one of two digital signals whose phases straddle the input phase +90°. If the integrator state is > 0, the chosen phase causes a net negative value to be integrated during that period. Conversely, if the integrator state is < 0, a net positive value will be integrated during that period.

2.3.2.1 – Cosine Non-Linearity

PDΣΔMs use choppers to detect the phases of near-sinusoidal signals and thus suffer from cosine non-linearity [23]. This is a result of the value that is accumulated in the integrator each ΣΔM-cycle, which for sinusoidal driving signals can be expressed as:

$$\Delta C = \frac{2}{180} \cdot \frac{A_f}{F_{drive}} \cdot \cos \Delta\varphi \qquad (2.24),$$

where ΔC is the number of counts integrated each ΣΔM-cycle, A_f is the frequency amplitude (which is half the peak-to-peak frequency deviation), F_{drive} is the driving frequency, which is equal to the ΣΔM frequency, and $\Delta\varphi$ is the difference between the input phase and the demodulating phase minus 90°. Because of the cosine term, this results in a non-linear bitstream average μ, which is plotted in Figure 2.17 for ±22.5° phase references. The error when linear bitstream averaging is applied is shown on the right and corresponds to ±2.5°C.

The disadvantage of the non-linear readout is that it requires higher-order polynomials to correct for this systematic non-linearity. Moreover, because of its distortion, it is more difficult to trim errors in the resistor. More importantly, it prevents the use of a multi-bit ΣΔM. Sine-wave

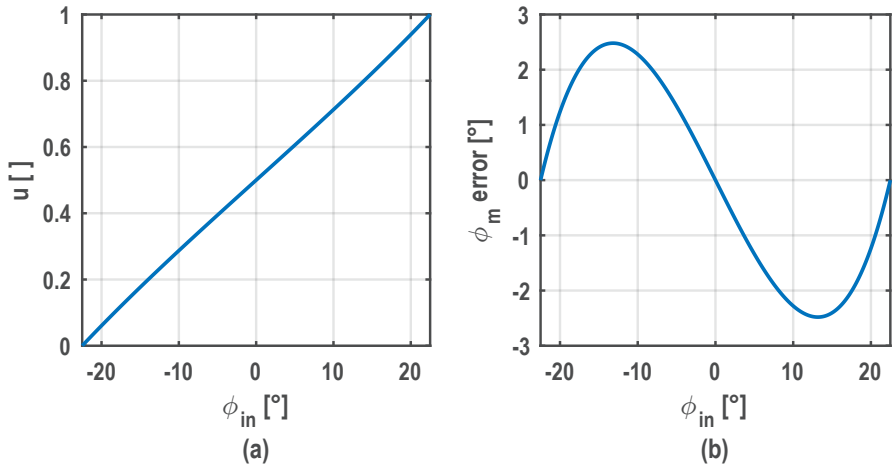

Figure 2.17: Cosine non-linearity in the bitstream average (a) and its phase error (b) for phase references of ±22.5°.

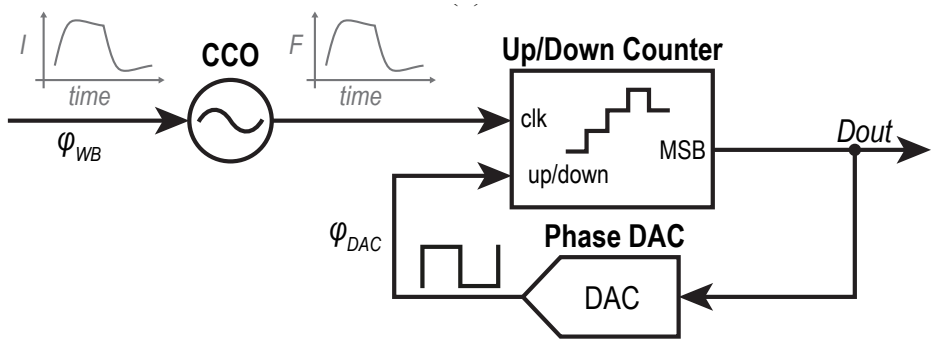

Figure 2.18: Block diagram of the highly digital PDΣΔM.

demodulation causes the cosine nonlinearity error seen in Figure 2.17 (b) in each LSB segment. It is virtually impossible to trim the sensor and correct for this error, even with higher-order polynomials, which require higher-order trimming.

2.3.2.2 – Highly Digital Implementation

The straightforward implementation of the PDΣΔM employs an analog integrator which requires capacitors and an amplifier. Consequently, this implementation is not area-efficient and requires headroom. As an alternative, the analog integrator can be replaced by a time-based integrator, as shown in Figure 2.18. The modulator's loop filter consists of a digital counter which is driven by a current-controlled oscillator (CCO) [21]. The latter converts the input signal from the current domain to the frequency

domain, allowing the counter to be used as an integrator. By toggling the counter's up/down signal, the polarity of this integration can also be toggled, effectively multiplying the input signal with the up/down signal. The modulator's 1-bit quantizer can then be realized by merely sampling and evaluating the counter's MSB. The result is a highly digital PDΣΔM [20], which can be easily scaled.

In the case of the highly digital PDΣΔM, the value that is integrated each ΣΔM cycle (ΔC) can be expressed as:

$$\begin{aligned}\Delta C &= \int_0^{T_{drive}/2} \left[0.5F_{pp}\sin(2\pi f_{drive}t + \varphi_{in}) + F_{os}\right]dt \\ &\quad - \int_{T_{drive}/2}^{T_{drive}} \left[0.5F_{pp}\sin(2\pi f_{drive}t + \varphi_{in}) + F_{os}\right]dt \\ &= \frac{F_{pp}\cos(\varphi_{in})}{\pi f_{drive}} \end{aligned} \quad (2.25),$$

where T_{drive} is the ΣΔM cycle period, which is equal to the driving period of the filter whose phase shift is digitized, F_{pp} is the CCO peak-to-peak frequency swing, f_{drive} is the ΣΔM cycle frequency, φ_{in} is the input phase (i.e., filter's output phase) that is digitized, and F_{os} is the CCO offset frequency. In the case of the counter based PDΣΔM, the unit of this value is in counts. This equation is the main equation used to determine phase sensitivity, and will be used many times throughout this thesis. In later chapters (Ch. 4 & 5) this equation is adjusted to reflect changes in the PDΣΔM architecture.

2.3.2.3 – Counter Wrap-Around

Unlike an analog integrator, a counter does not clip. Instead, it wraps around. This corrupts the counter's state and must be prevented. After one ΣΔM period, the number of counts accumulated by the counter can be simplified to [22]:

$$C_{\Delta} \approx \frac{F_{CCO,pp}}{180° \cdot F_{drive}} \cdot (\varphi_e) \quad (2.26),$$

where $F_{CCO,pp}$ is the peak-to-peak CCO frequency variation, F_{drive} is the filter's driving frequency, which is equal to the ΣΔM sampling frequency.

In a PDΣΔM, the counter's state oscillates around its mid-code, so to prevent wrap-around, C_{Δ} should never exceed half the counter length. As a result, the counter's size is determined by the CCO frequency swing, the ΣΔM period, and the maximum input phase error. Consequently, the power consumption is mostly determined by the CCO frequency. A higher CCO frequency swing requires a higher average CCO frequency, which causes higher power consumption in the counter. Moreover, it also increases the counter size requirements, resulting in even higher power consumption. Consequently, in these highly digital designs the counter is the main power consumer.

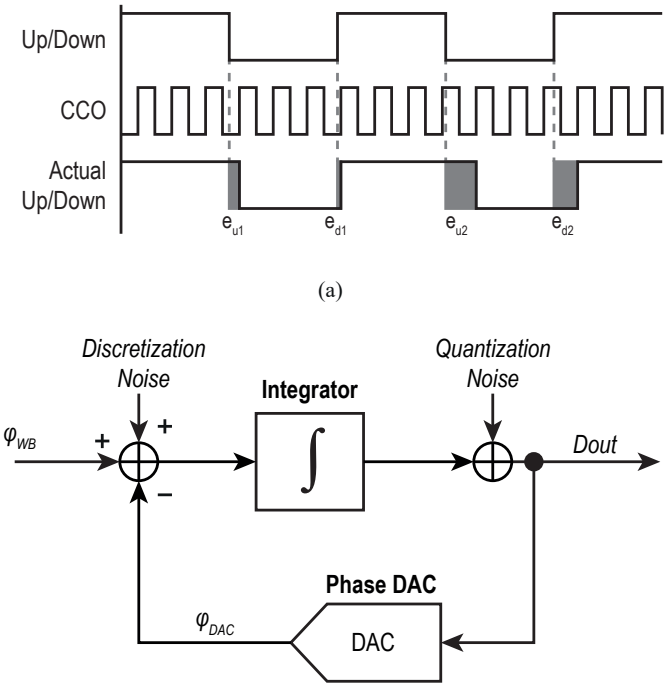

Figure 2.19: (a) Timing diagram of the up/down and CCO signals, and (b) Linear model of the VCO-based PDΣΔM.

2.3.2.4 – Time-Quantization Noise

The downside of using a counter as an integrator is that it quantizes the phase of the CCO, and therefore introduces time-domain quantization noise [22]. Figure 2.19 (a) illustrates this: when the up/down signal toggles, the CCO phase, and therefore the integrator's state, is quantized. In this example, the CCO experiences approximately 3.15 cycles in the first "up" counting period. However, the counter truncates this to three cycles, creating an error in the integrator state. On top of that the next cycle starts at the truncated value, creating an error in the subsequent "down" counting period. As shown in Figure 2.19 (b), this time-quantization (TQ) noise can be modeled as an additive white-noise source at the input of the integrator [20], and so does not benefit from noise shaping.

To calculate the in-band TQ noise, the integrated error per ΣΔM-period is calculated. Each ΣΔM-period contains two moments at which the up/down signal is toggled. During these transitions, the phase of the CCO

is quantized, which incurs a uniform error between 0 and 1 CCO-period (i.e., between 0 and T_{CCO}). The resulting integration error is equal to the time shift ($1/\sqrt{12} \cdot T_{CCO}$) multiplied by the input jump ($2 \cdot F_{CCO}$). There are two chopping transitions in each ΣΔM-period, so the integrated error each period is:

$$\sigma_{\Delta C}{}^2 = 2 \cdot \left(\frac{1}{\sqrt{12}} \cdot T_{CCO} \cdot 2 \cdot f_{CCO}\right)^2 = \frac{2}{3} \quad (2.27).$$

The resulting in-band phase noise is then:

$$\sigma_{P,IB} = \sqrt{\frac{T_{drive}}{T_{conv}}} \cdot \sigma_{\Delta C} \cdot \frac{180° \cdot F_{drive}}{F_{cco,pp}} = \sqrt{\frac{\sigma_{\Delta C}{}^2}{OSR}} \cdot \frac{180° \cdot F_{drive}}{F_{cco,pp}} \quad (2.30),$$

where T_{drive} is the ΣΔM and driving period, T_{conv} is the total ΣΔM conversion time, and the OSR is the oversampling ratio. Assuming that the transitions of the up/down signal are random with respect to the CCO phase—which is the case if the CCO frequency (F_{CCO}) is asynchronous with respect to the demodulating frequency (F_{drive}), and/or if F_{CCO} is dithered by thermal noise—the counter's rounding error will have a uniform distribution. In that case, the total in-band phase noise (in radians) is given by [22]:

$$\sigma_{P,IB} = \sqrt{\frac{2}{3 \cdot OSR}} \cdot \frac{\pi \cdot F_{drive}}{F_{CCO,pp}} \quad (2.31).$$

It can be seen that the discretization noise can be minimized either by decreasing F_{drive}, increasing the modulator's OSR, or increasing the CCO's frequency swing ($F_{CCO,pp}$). Decreasing the driving frequency comes at the cost of area, since the filter's corner frequency must also decrease. Increasing the CCO frequency swing comes at the cost of increased counter power consumption, since it runs at a higher frequency and the counter size has to increase to prevent wrap-around. In practice, counter-based PDΣΔMs are much less energy-efficient than their analog counterparts. However, their area efficiency is also superior, since the integrator size (a small counter) is

Table 2.2: Overview of variations of important PDΣΔM metrics.

	CCO + w/o windowing (Ch. 3)	Fixed OSC (digital mixer) (Ch. 4)	Fixed OSC (digital mixer) + windowing (Ch. 4)	GRO (digital mixer) (Ch. 5)
Integrated value	$\frac{F_{cco,pp}}{180° \cdot F_{drive}} \cdot \cos(\varphi_e)$	$\frac{2 \cdot F_{OSC}}{180° \cdot F_{drive}} \cdot \varphi_e$	$\frac{2 \cdot F_{OSC}}{180° \cdot F_{drive}} \cdot \varphi_e$	$\frac{2 \cdot F_{GRO}}{180° \cdot F_{drive}} \cdot \varphi_e$
Per period TQ error	$\frac{2}{3}$	$\frac{4}{3}$	$\frac{3}{3}$	—
Per conversion TQ error	—	—	—	$\frac{4}{2}$
In band phase-noise	$\frac{90^\circ}{F_{OSC}} \cdot \sqrt{\frac{F_{drive}}{T_{conv}} \cdot \frac{4}{3}}$	$\frac{90^\circ}{F_{OSC}} \cdot \sqrt{\frac{F_{drive}}{T_{conv}} \cdot \frac{3}{3}}$	$\frac{90^\circ}{F_{GRO}} \cdot \sqrt{\frac{F_{ChL}}{T_{conv}} \cdot \frac{4}{2}}$	$\frac{180°}{F_{cco,pp}} \cdot \sqrt{\frac{F_{drive}}{T_{conv}} \cdot \frac{2}{3}}$

much smaller than a capacitor-based integrator. In thermal management, the resolution requirement is very relaxed, but the area efficiency is very strict.

In later chapters, the PDΣΔM architecture is modified for improved performance which introduces differences in the ΣΔM-cycle integrated value (ΔC), the TQ noise, and the in-band phase noise. Table 2.2 gives an overview of these differences.

2.4: References

[1] D. A. Neamen, Semiconductor Physics and Devices - Basic Principles, Third ed., McGraw-Hill, 2003.

[2] S. Sze, Semiconductor Devices: Physics and Technology, John wiley & sons, 2008.

[3] C. H-M and *et al.*, "A comprehensive study of polysilicon resistors for CMOS ULSI applications," *Superlattices and Microstructures*, vol. 33, no. 4, pp. 193-208, 2003.

[4] I.-T. Lu, J.-J. Zhou, J. Park and M. Bernardi, "First-principles ionized-impurity scattering and charge transport in doped materials," *Physical Review Materials*, vol. 6, no. 1, 2022.

[5] S. X. Shian and D. Z. Pan, "Wire sizing with scattering effect for nanoscale interconnection," in *Asia and South Pacific Conference on Design Automation*, Yokohama, 2006.

[6] J. Park, J. Kim, K. Kim, J.-H. Yang, M. Choi, and J. Shin, "A 0.65V 1316μm² Fully Synthesizable Digital Temperature Sensor Using Wire Metal Achieving 0.16nJ·%²-Accuracy FoM in 5nm FinFET CMOS," in *International Solid-State Circuit Conference*, San Fransisco, 2022.

[7] W. A. Lane and G. T. Wrixon, "The design of thin-film polysilicon resistors for analog IC applications," *IEEE Transactions on Electron Devices*, vol. 36, no. 4, pp. 738-744, 1989.

[8] S. Pan, Y. Cheng, G. Wu, Z. Wang, K. A. A. Makinwa, and H. Wu, "A 0.028mm² 32MHz RC Frequency Reference in 0.18μm CMOS with ±900ppm Inaccuracy from −40°C to 125°C and ±1600ppm Inaccuracy After Accelerated Aging," in *International Solid-State Circuits Conference*, San Francisco, 2024.

[9] F. Nava, K. Tu, O. Thomas, J. Senateur, R. Madar, A. Borghesi, G. Guizzetti, U. Gottlieb, O. Laborde, and O. Bisi, "Electrical and optical properties of silicide single crystals and thin films," *Materials sience reports*, vol. 9, no. 4-5, pp. 141-200, 1993.

[10] R. H. Bube, "Electronic transport in polycrystalline films," *Annual Review of Materials Science*, vol. 5, no. 1, pp. 201-224, 1975.

[11] T. Ohguro, S. Nakamura, M. Koike, T. Morimoto, A. Nishiyama, Y. Ushiku, T. Yoshitomi, M. Ono, M. Saito, and H. Iwai, "Analysis of resistance behavior in Ti- and Ni-salicided polysilicon films," *IEEE Transaction on Electron Devices*, vol. 41, no. 12, pp. 2305-2317, December 1994.

[12] K-S. Park, N. Pal, Y. Li, R. Xia, T. Wang, A. Abdelrahman, and P. K. Hanumolu, "A Temperature- and Aging-Compensated RC Oscillator With ±1030-ppm Inaccuracy From -40 °C to 85 °C After Accelerated Aging for 500 h at 125 °C," *IEEE Journal of Solid-State Circuits*, vol. 58, no. 12, pp. 3459-3469, Dec. 2023.

[13] S. Pan and K. A. A. Makinwa, "A 10 fJ·K² Wheatstone Bridge Temperature Sensor With a Tail-Resistor-Linearized OTA," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 2, pp. 501-510, 2021.

[14] M. Perrott, J. Salvia, F. Lee, A. Partridge, S. Mukherjee, C. Arft, J.-T. Kim, N. Arumugam, P. Gupta, S. Tabatabaei, S. Pamarti, H. Lee, and F. Assaderaghi, "A Temperature-to-Digital Converter for a MEMS-Based Programmable Oscillator with Better Than ±0.5ppm Frequency Stability," in *IEEE International Solid-State Circuits Conferenc*e, San Francisco, 2012.

[15] A. Khashaba, J. Zhu, A. Elmallah, M. Ahmed, and P. K. Hanumolu, "A 0.0088mm² Resistor-Based Temperature Sensor Achieving 92fJ·K² FoM in 65nm CMOS," in *IEEE International Solid-State Circuits Conference*, 2020.

[16] S. Pan, Ç. Gürleyük, M. F. Pimenta, and K. A. A. Makinwa, "A 0.12mm² Wien-Bridge Temperature Sensor with 0.1°C (3σ) Inaccuracy from -40°C to 180°C," in *IEEE International Colid-State Circuits Conference*, 2019.

[17] C. P. van Vroonhoven and K. A. A. Makinwa, "Thermal diffusivity sensors for wide-range temperature sensing," in *SENSORS*, Lecce, Italy, 2008.

[18] A. Mordakhay and J. Shor, "Miniaturized, 0.01 mm², Resistor-Based Thermal Sensor With an Energy Consumption of 0.9 nJ and a Conversion Time of 80 µs for Processor Applications," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 10, pp. 2958-2969, 2018.

[19] W. Choi, Y. Lee, S. Kim, S. Lee, J. Jang, J. Chun, K. A. A. Makinwa, and Y. Chae, "A Compact Resistor-Based CMOS Temperature Sensor With an Inaccuracy of 0.12 °C (3σ) and a Resolution FoM of 0.43 pJ · K² in 65-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 12, pp. 3356-3367, 2018.

[20] Y. Lee, W. Choi, T. Kim, S. Song, K. A. A. Makinwa, and Y. Chae, "A 5800- µm² Resistor-Based Temperature Sensor With a One-Point Trimmed Inaccuracy of ±1.2 °C (3σ) From −50 °C to 105 °C in 65-nm CMOS," *IEEE Solid-State Circuits Letters*, vol. 2, no. 9, pp. 67-70, 2019.

[21] R. Quan, U. Sonmez, F. Sebastiano, and K. A. A. Makinwa, "A 4600μm² 1.5°C (3σ) 0.9kS/s thermal-diffusivity temperature sensor with VCO-based readout," in *IEEE International Solid-State Circuits Conference*, San Fransisco, 2015.

[22] U. Sönmez, F. Sebastiano, and K. A. A. Makinwa, "Analysis and Design of VCO-Based Phase-Domain ΣΔ Modulators," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 5, pp. 1075-1084, 2017.

[23] S. M. Kashmiri, S. Xia, and K. A. A. Makinwa, "A Temperature-to-Digital Converter Based on an Optimized Electrothermal Filter," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 7, pp. 2026-2035, 2009.

Chapter 3 | Design I: Analog Readout & Limited Accuracy and Area

This chapter describes a first attempt to design a compact resistor-based temperature sensor. At this time, no sub-10,000 μm² resistor-based temperature sensors had ever been published. To do this, the architecture of a previous design [1] was adapted as follows: 1) its differential Wien-Bridge (WB) filter was replaced by a smaller single-ended implementation and 2) instead of a highly analog phase-domain delta-sigma modulator (PDΣΔM) [2], it employs the highly digital, and thus highly scalable design reported in [3].[1]

[1] This chapter is based on [11].

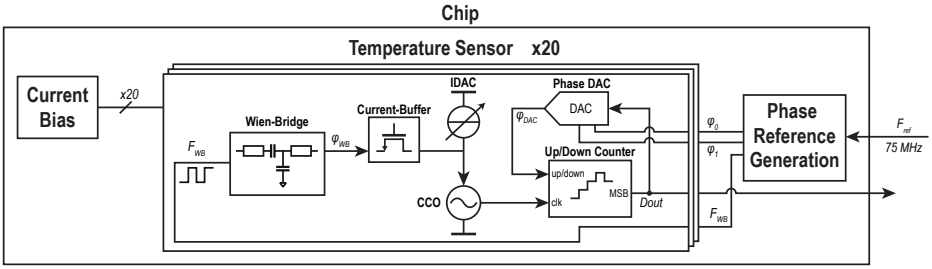

Figure 3.1: Simplified system level block-diagram.

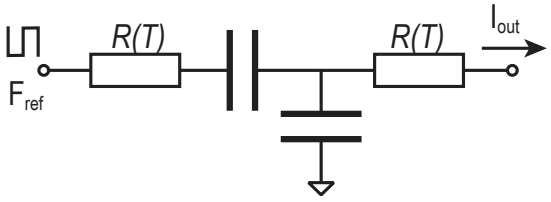

Figure 3.2: Miniaturized single-ended Wien-Bridge.

Fig. 3.1 shows a simplified block diagram of the proposed system. It incorporates 20 sensors on a single die, which facilitates rapid characterization. They share a bias current- and a phase reference generator (section 3.5), the latter being driven by an external frequency-reference F_{ref} (= 75 MHz). Each sensor consists of a single-ended WB (section 3.1), whose phase-shifted output current is digitized by a PDΣΔM [2] (section 3.2). Instead of using an analog integrator, which requires large capacitors [2], a compact counter-based integrator is used. The counter is driven by a current-controlled oscillator (CCO) [3] (section 3.3), which in turn is driven by the WB filter via a current buffer (section 3.4). To provide robustness against process spread, the center frequency of the CCO can be trimmed via a current DAC (IDAC). In the following sections, the design of each of these sub-blocks and their impact on sensor performance will be discussed in more detail.

3.1: Single-Ended Wien-Bridge

As discussed above, a single-ended WB is used (Fig. 3.2) because it requires fewer components and area than a differential implementation [1]. This is enabled by the fact that being a band-pass filter, a WB can have different input and output DC-levels. The WB is realized with silicided p-poly resistors (R_{WB} = 28 kΩ) and MiM capacitors (C_{WB} = 1.84 pF), resulting in a

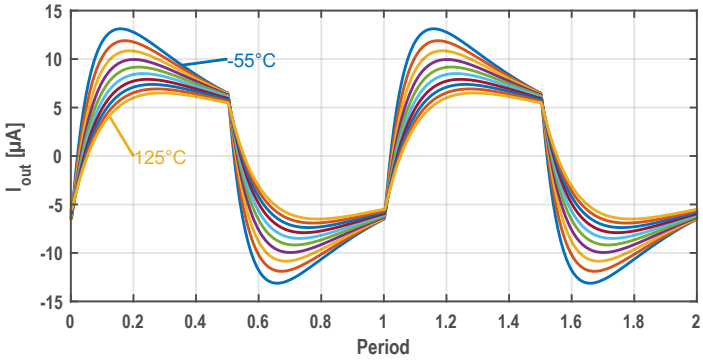

Figure 3.3: WB transient output current for different temperatures.

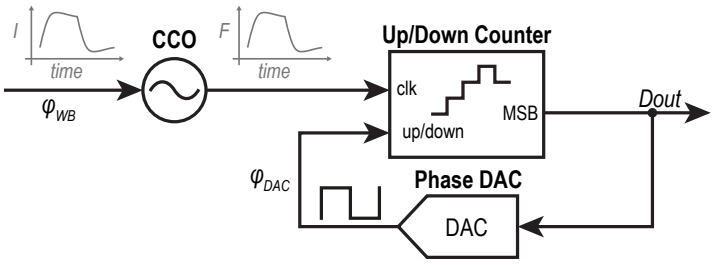

Figure 3.4: CCO-based Phase-Domain Sigma-Delta Modulator[3].

center frequency F_{WB} = 3 MHz at room temperature. In the chosen 180 nm process, it occupies only 3700 µm². When driven by a 1.8 V square-wave at F_{WB}, however, the WB's output voltage swing will exceed the linear range of a simple differential pair. For this reason, its output current is usually read out by connecting the resistors in its output branch to the virtual ground of an active integrator [1] or to a current buffer [4]. The resulting output current is shown in Fig. 3.3. It can be seen that both its shape (phase) and amplitude vary significantly over temperature.

3.2: Phase-Domain Simga-Delta Modulator

As shown in Fig. 3.4, the modulator's loop filter consists of a digital counter, which is driven by a CCO [3]. The latter converts the input signal from the current domain to the frequency domain, allowing a counter to be used as an integrator. By toggling the counter's up/down signal, the polarity of this integration can also be toggled, effectively multiplying the input signal with the up/down signal. The modulator's 1-bit quantizer can then be realized by merely sampling and evaluating the counter's MSB. The result is a highly-digital PDΣΔM [3] (Ch. 2.3.2.2), which can be easily scaled.

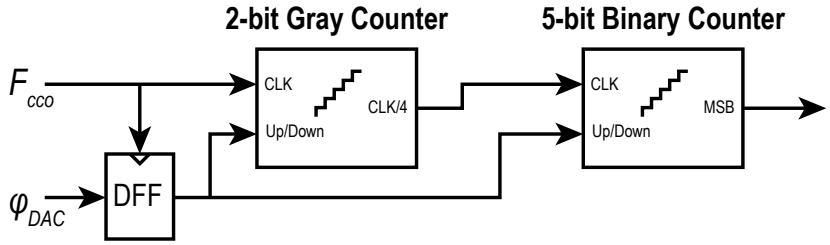

Figure 3.5: 7-bit counter/integrator. Split into a 2-bit Gray-code counter and a coarse 5-bit binary counter [6].

3.2.1 – Counter Time-Quantization Noise

The time discrete nature of the counter/integrator in the highly-digital PDΣΔM causes time-quantization (TQ) noise (Ch. 2.3.2.4) [5]. This can be minimized by maximizing the CCO's frequency swing, but this comes at the cost of power consumption. In this design, F_{WB} = 3 MHz, which in turn limits the modulator's sampling frequency to 3 MHz, since $F_s \leq F_{WB}$ in a PDΣΔM. The OSR is then set by the signal bandwidth, which is defined by the desired conversion rate (3 kSa/s) and the choice of a simple sinc[1] decimation filter. A large CCO frequency swing requires an even larger CCO center frequency, and hence, high power consumption. In this design $F_{CCO,pp}$ is set to 600 MHz, which results in a TQ-noise limited resolution of 0.12°C (rms). Although this is much worse than the mK-level resolution of the WB itself [4], it is good enough for thermal monitoring applications.

3.2.2 – Up/down Counter

Compared to an analog integrator, instead of clipping, a counter wraps around (Ch. 2.3.2.3). This corrupts the counter's state and must be prevented. From simulations, φ_{WB} varies by about 23° over the targeted temperature range (-35°C—125°C). With ±40 % spread in RC, this results in a worst-case phase difference $\varphi_{WB} - \varphi_{DAC}$ of about 40°. Given $F_{CCO,pp}$ = 600 MHz and F_{WB} = 3 MHz, Eq. (2.26) indicates that the maximum value integrated in the counter during each ΣΔM-cycle (C_Δ) is 44 counts. In the worst-case, this may occur twice in succession and so a 7-bit counter is required to prevent wrap-around.

To guarantee a 600 MHz (pp) frequency swing even in the presence of process variations, both the CCO and the counter are designed to operate up to 800 MHz over PVT. To achieve this, the counter is split up into two

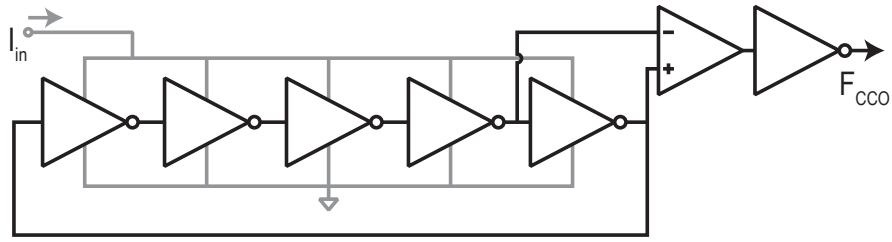

Figure 3.6: 5-stage ring-oscillator with a differential-pair to level-shift the CCO swing to a full digital waveform.

parts: a fine two-bit counter and a coarse five-bit counter (Fig. 3.5) [6]. This reduces its power consumption, since clock-gating can then be used to ensure that the coarse counter only operates at one quarter the frequency of the fine counter. The fine-counter is a gray-code counter, resulting in a faster and simpler implementation. To limit meta-stability issues to a single D-FF rather than to the entire counter, the up/down signal φ_{DAC} is re-clocked by F_{CCO}, before being applied to the counter.

The counter was synthesized using the normal digital design flow of the 180 nm CMOS process. Operating at 800 MHz, it consumes 1 mW from a 1.8 V power supply, making it the most power-hungry sub-block of the sensor. However, in a more advanced process, its power consumption will scale dramatically, e.g., to about 130 µW in a 65 nm process.

3.3: CCO

The CCO consists of a 5-stage chain of inverters, which converts the output current of the WB (30 µA pp) into the desired 600 MHz frequency swing. To compensate for the CCO center frequency's sensitivity to PVT, a 6-bit current DAC (~30 MHz LSB) is used to trim the CCO's center frequency to ~400 MHz at room temperature. Since this frequency is too high to be readily measured off-chip, the up/down counter is used as a divide-by-128 counter during trimming.

Since the inverter stages of the CCO do not output logic-compatible signals, their outputs are amplified and level-shifted by a differential-pair (Fig. 3.6). After this, a simple inverter is enough to generate a rail-to-rail logic swing.

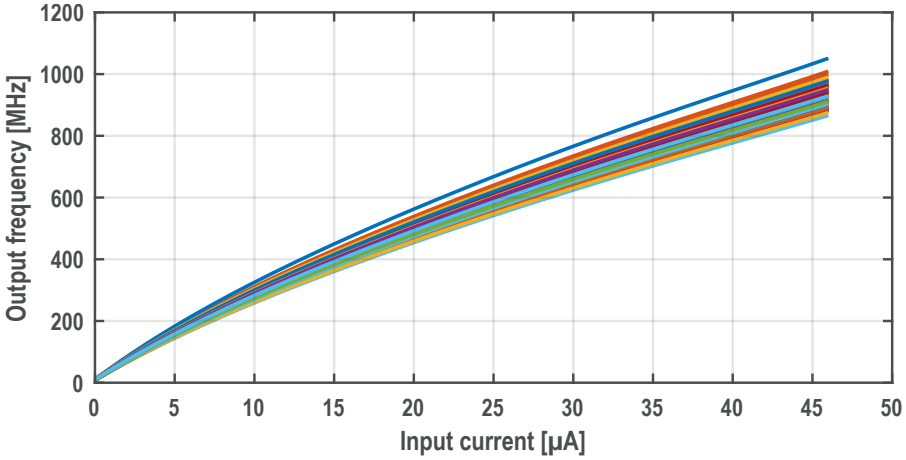

Figure 3.7: Monte-Carlo (process + mismatch) simulation of the CCO current-to-frequency transfer.

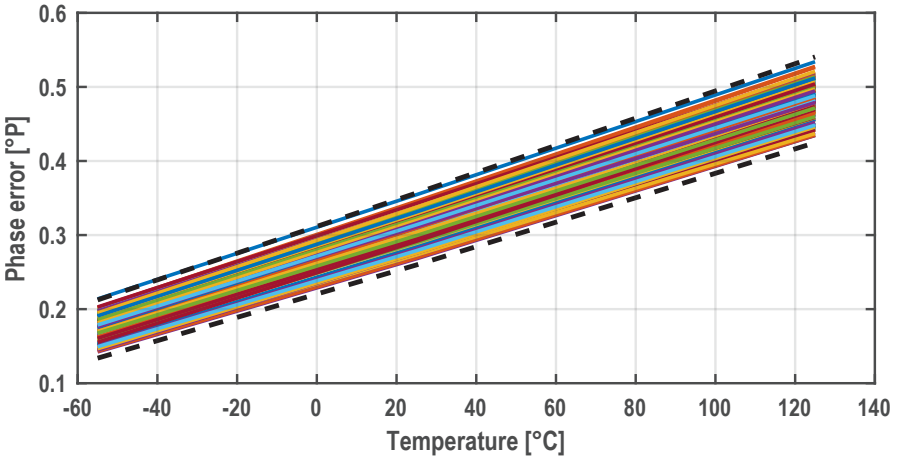

Figure 3.8: Simulated additional phase-shift due to the CCO non-linearity at -55, 25, and 125°C (Monte-Carlo process + mismatch). The dashed lines indicate the 3σ values.

3.3.1 – CCO Non-Linearity

As shown in Fig. 3.7, the CCO's current to frequency characteristic is quite non-linear. Since its output is multiplied by the square-wave output of the phase DAC, distortion components at odd harmonics of F_{WB} will fold back to base-band. In other words, CCO non-linearity will cause errors in the PDΣΔM's error signal φ_e and, hence, in its bitstream average. Although CCO linearity could be improved by limiting its frequency swing, this

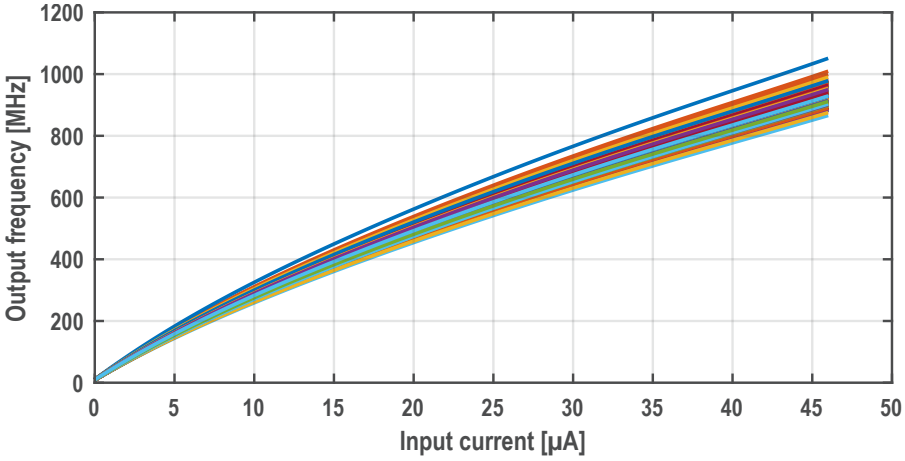

Figure 3.9: Simulator Monte-Carlo (process + mismatch) spread due to the amplitude variations in the WB output current.

would increase the counter's discretization noise. Fortunately, CCO non-linearity will result in systematic errors in the sensor's transfer function, which can then be accounted for during calibration, while any spread can be mitigated by trimming.

Monte-Carlo simulation results are plotted in Fig. 3.8, and show the additional phase-shift resulting from variations in CCO non-linearity. Although this is significant (ranging from 0.2° to 0.5°), its spread (mainly offset) is relatively small, and can be reduced to less than 80 mK (3σ) after a 1-point trim.

Variations in the WB's output current swing will also vary the effect of CCO non-linearity, which, in turn, will cause variations in the detected WB phase. This may be due to power supply variations, and/or the spread (±20%) of the WB resistors. The former will result in a finite power supply sensitivity, which is simulated to be 5.3°C/V, and is close to the measured supply sensitivity of 4.6°C/V. The latter will cause variations in the measured phase (Fig. 3.9) and therefore increase inaccuracy. This can only be somewhat mitigated by trimming: to 1.2°C (3σ) after a 1-point trim, and to 0.13°C (3σ) after a 2-point trim.

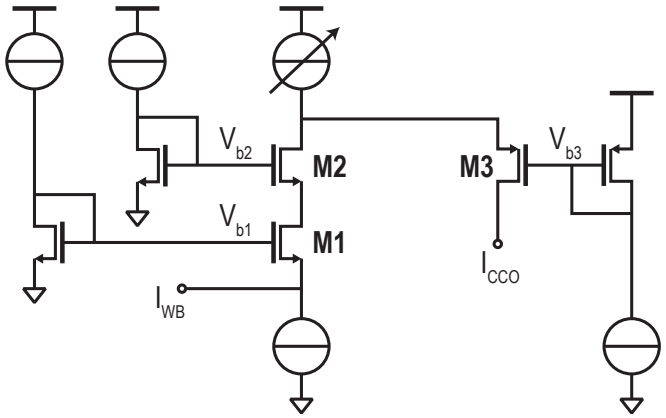

Figure 3.10: Circuit diagram of the current-buffer. The cascodes in the current-buffer reduce the input impedance as seen from the Wien-Bridge side.

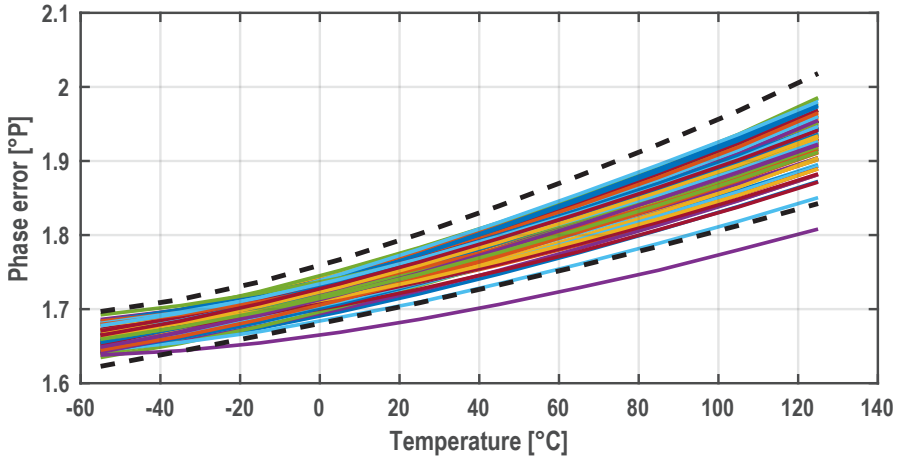

Figure 3.11: Simulated Monte-Carlo (process + mismatch) spread due to current buffer and bias spread.

3.4: Current Buffer

As seen from its supply terminals, the impedance of the CCO is quite significant compared to that of the WB resistors. To avoid altering its phase response, a current buffer is inserted between the WB and the CCO. As shown in Fig. 3.10, the current buffer consists of a common-gate amplifier (M1), with an input impedance ~ 600Ω << R_{WB} at F_{WB}, and a folded-cascode output that drives the CCO. Since they are both small (for speed and area)

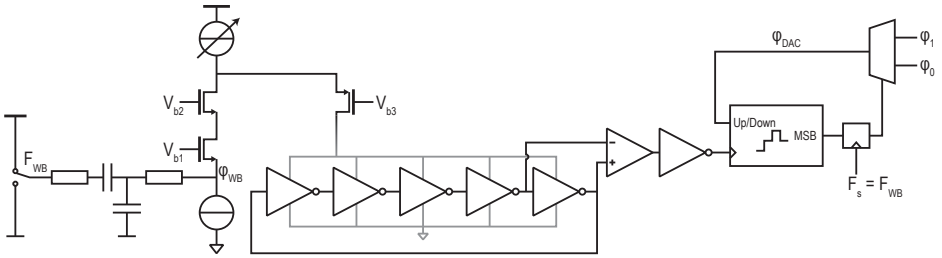

Figure 3.12: Complete sensor overview.

and the CCO input has a large voltage swing, two cascodes (M2, M3) are used to isolate the voltage swing across the CCO from the WB. The required bias voltages are generated by diode-connected MOSFETs, each of which is biased with a nominal 5.6 µA.

Fig. 3.11 shows the simulated phase-shift spread due to the spread of the current buffer and its bias current. Once more, there is a relatively large average shift (~1.8°) and a non-linear temperature dependency, both of which will be corrected during the sensor's calibration. Simulations show that the resulting temperature error is less than 0.6°C after a 1-point, and less than 0.1°C after a 2-point trim.

3.5: Shared Bias and Phase Reference Generation

Fig. 3.12 shows the complete system, in which the WB is driven by a square-wave signal at F_{WB}, which then generates an output current with a phase-shift φ_{WB}. This is then fed to the current buffer and used to drive the CCO, which, in turn drives the PDΣΔM. A phase generator provides the required reference signals (F_{WB} = 0°, φ_0 = 90° - 20° and φ_1 = 90° + 20°).

The phase reference generator is driven by an external 75 MHz clock. It consists of a programmable divide-by-2N counter, whose output is a square-wave with a frequency of F_{WB} and a 50% duty-cycle. This is then used to drive the WB, as well as a programmable delay chain, which generates the phase references used by the phase DAC. As shown in Fig. 3, the phase generator is shared by all 20 temperature sensors on the same chip. It occupies 1200 µm² and dissipates less than 0.1 mW (simulated). It should be noted that a non-programmable generator would only have occupied 430 µm².

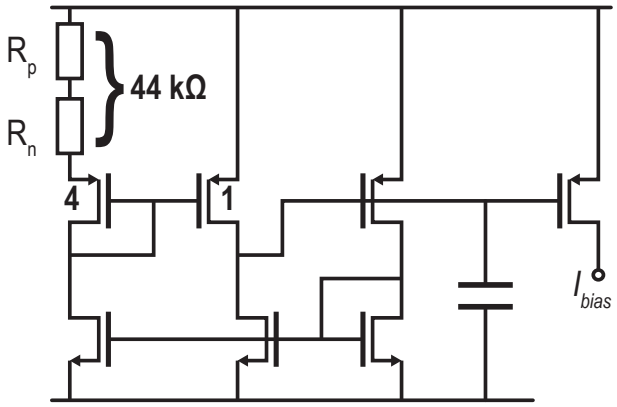

Figure 3.13: Circuit diagram of the bias source.

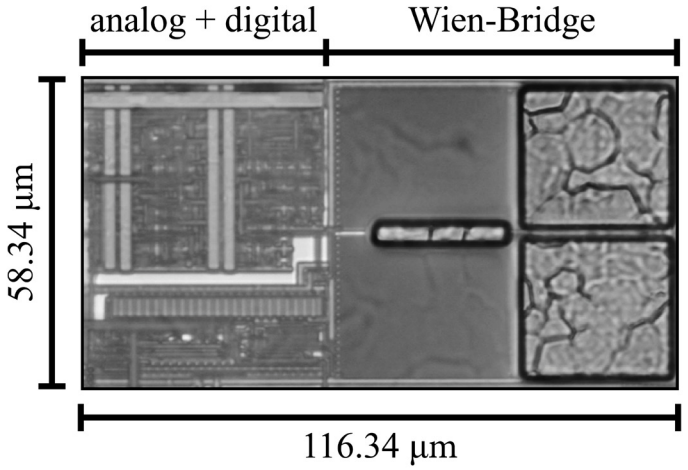

Figure 3.14: Chip photograph.

The bias current generator (Fig. 3.13) operates by forcing a PTAT voltage across a resistor. A composite resistor, made by combining two resistors with opposite temperature coefficients, ensures that the resulting current is relatively flat over temperature. This current is then mirrored out to all 20 sensors. The residual curvature is less than 10%, which means that a single trim is sufficient to ensure that the CCO operates properly over the military temperature range. The bias current generator dissipates about 33 µW and occupies 2500 µm² (however, it was not designed for area efficiency).

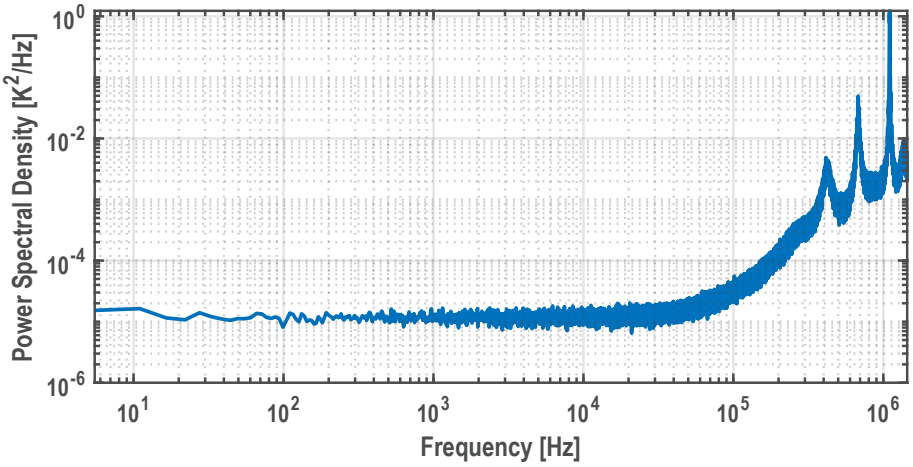

Figure 3.15: PSD of 2^{19} bit-stream samples (with the mean value removed to better observe LF noise), averaged 100x.

3.6: Measurements

The system was realized in a standard 180-nm CMOS technology (Fig. 3.14). Each chip contains 20 temperature sensors, which each occupy 6800 µm², as well as a shared bias current and a phase reference generator. The system operates from a 1.8 V supply voltage and dissipates 1.6 mW (60% is consumed by the counter, 39% by the current buffer and CCO, and 1% by the WB).

Due to the finite input impedance of the current buffer, and the spread and parasitic capacitances of the WB resistors, the average center frequency of the implemented WB filters is a bit lower than 3 MHz. To accommodate this, all the sensors were driven at $F_{WB} \approx 2.9$ MHz (= 75 MHz / 2·13). An FFT of the PDΣΔM's bitstream output (Fig. 3.15), confirms that the sensor's noise floor is indeed dominated by the counter's discretization noise. After decimation by an off-chip 1024-tap sinc¹ filter, the sensor achieves a resolution of 0.12°C (1σ) at a conversion rate of 2.8 kSa/s, which is in good agreement with the discretization noise levels predicted by Eq. (2.31).

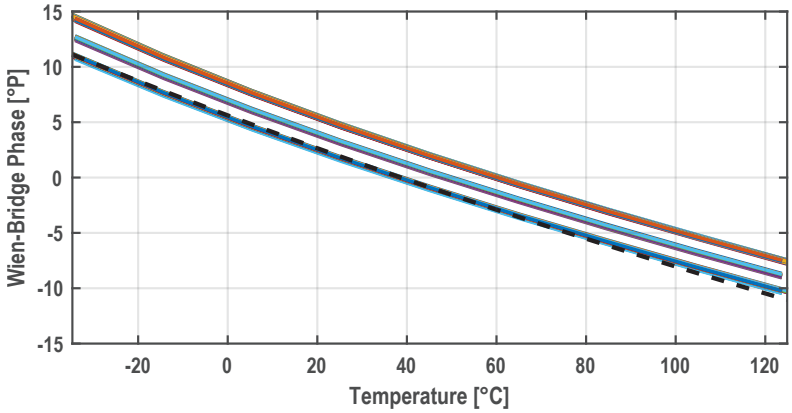

Figure 3.16: Measured phase of 160 samples from 8 chips compared to simulated phase (dashed line).

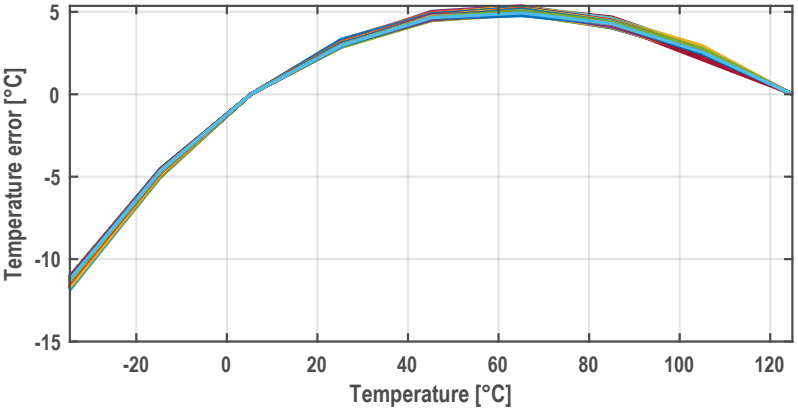

Figure 3.17: Temperature error of 160 samples from 8 chips after 2-pt trim, but before systematic non-linearity correction.

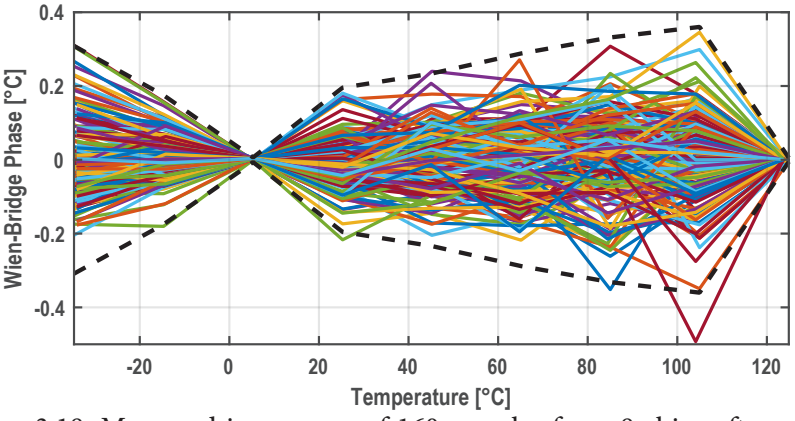

Figure 3.18: Measured inaccuracy of 160 samples from 8 chips after a 2-pt temperature trim (dashed line indicates the 3σ error).

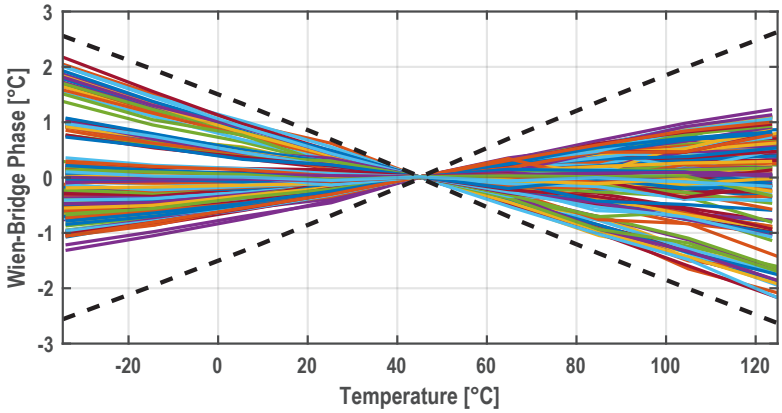

Figure 3.19: Measured inaccuracy of 160 samples from 8 chips after a 1-pt temperature trim (dashed line indicates the 3σ error).

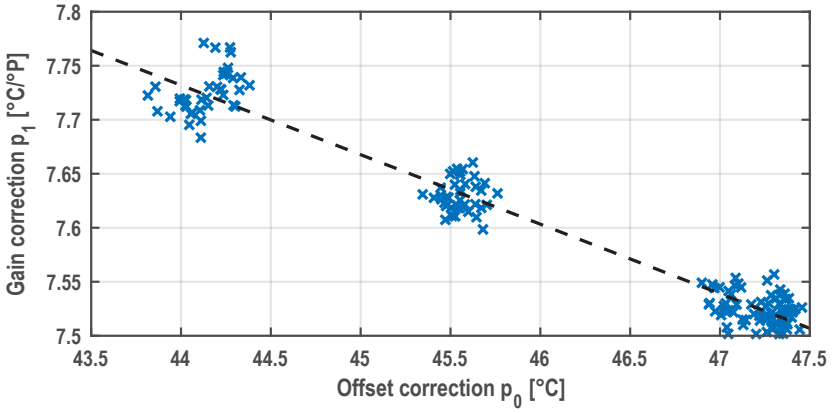

Figure 3.20: Correlation between offset correction p_0 and gain correction p_1.

The sensor's output over temperature is shown in Fig. 3.16 (160 samples from 8 chips). After a 2-point trim (at 5°C and 125°C), the remaining error is dominated by a systematic non-linearity (Fig. 3.17). This can be removed by a fixed 3rd order polynomial [4], resulting in an inaccuracy of ±0.35°C (3σ) from -35°C to 125°C (Fig. 3.18).

After a 1-point offset trim, the sensor achieves an inaccuracy of ±2.7°C (3σ) (Fig. 3.19). This can be improved by noting that the offset and gain errors of individual sensors with respect to their average master curve are correlated (Fig. 3.20). Simulations show that this correlation is mainly due to the phase errors caused by CCO non-linearity. As a result, better accuracy can be obtained by a correlated 1-point trim [1], in which a sensor's gain

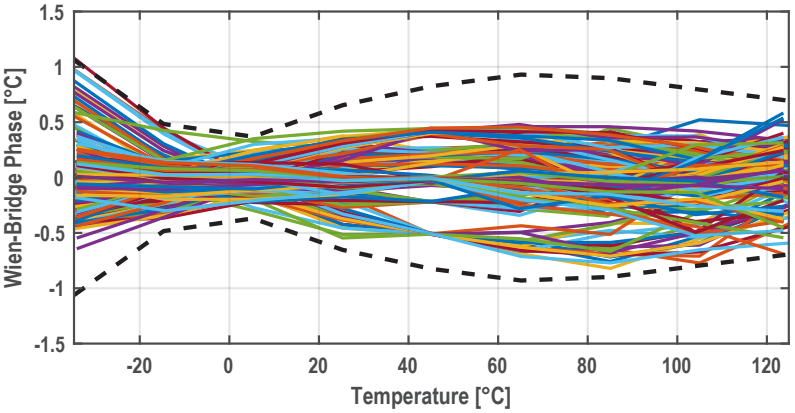

Figure 3.21: Measured inaccuracy of 160 samples from 8 chips after a correlated 1-pt temperature trim (dashed line indicates the 3σ error).

error is estimated from its offset at 45°C. This results in an inaccuracy of ±1.2°C (3σ) (Fig. 3.21). In Table 3.1, the sensor's performance is compared to other state-of-the-art temperature sensors. The area and power of the bias and phase generators are not included, since these are both negligible compared to that of the 20 sensors.

As stated before, the CCO center frequency will spread over PVT and so must be trimmed to about 400 MHz. Since the intra-batch CCO spread is small and the same bias current generator is used, the same trim code can be used for all the CCOs on one die. This means that in practice, only one CCO (of the 20) needs to be trimmed, which considerably reduces the required effort.

3.7: Conclusion

This chapter describes the design of a compact and accurate resistor-based temperature sensor. Compared to the state-of-the-art at that time, it achieves competitive area and accuracy, despite being realized in a mature 180 nm process. Its low area is mainly due to the use of a highly digital CCO-based ADC. Most of the design choices were made while keeping in mind the characteristics of the more advanced process nodes in which such sensors are typically used. As a result, this design scales well with technology and its performance is expected to improve when ported to more advanced processes.

Table 3.1: Performance Summary and Comparison with State-of-the-Art.

Publication	This Work	U. Sonmez [7]	S. Pan [1]	W. Choi [8]	Y-C Hsu [9]	M. Eberlein [10]
Year	2018	2016	2018	2018	2017	2016
Type	Resistor	TD	Resistor	Resistor	PNP	NPN
Technology	180 nm	40 nm	180 nm	65 nm	28 nm	28 nm
Area (μm²)	6 800[1]	1 650[1]	720 000[1]	7 000[1]	9 461	3 800
Inaccuracy (3σ, °C)	—	1.4	—	—	0.87	1.8
Inaccuracy, 1-pt trim (3σ, °C)	—	0.75	0.2	2.47	—	—
Inaccuracy, 2-pt trim (3σ, °C)	0.35	—	0.03	0.12	—	—
Temperature Range (°C)	-35 to 125	-40 to 125	-40 to 85	-40 to 85	0 to 125	-20 to 130
Resolution (°C)	0.12	0.36	0.00041	0.0025	0.15	0.5
Speed (kS/s)	3	1	0.2	1	0.12	31.25
Supply Voltage (V)	1.8	0.9 – 1.2	1.6 – 2.0	0.85 – 1.05	1.8	1.1 – 2.0
Supply Sensitivity (°C/V)	4.6	2.8	0.17	0.5	—	0.1[2]
Power (mW)	1.6[1]	2.5[1]	0.16[1]	0.068[1]	0.01875	0.0176

[1] Additional circuitry required (e.g., bias current and phase reference generators (this work), decimation filters (this work, [7, 1] or frequency-to-digital converters [8]). [2] Simulated, measured < 0.5 LSB ~ 0.33°C/V.

The design has two downsides: 1) the current-buffer limits its portability and scalability, and 2) the non-linear current-to-frequency transfer of the CCO complicates the trimming process and ultimately limits both its accuracy and PSS. Although the area of the current buffer is relatively small in this design, this is expected to change when it is ported to more advanced technologies. Moreover, it requires a bias current to be distributed across the die, which is undesirable when sensors must be embedded in dense digital circuitry. Therefore, the design described in the next chapter will focus on addressing these issues.

3.8: References

[1] S. Pan, Y. Luo, S. H. Shalmany, and K. A. A. Makinwa, "A Resistor-Based Temperature Sensor With a 0.13 pJ·K² Resolution FoM," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 1, pp. 164-173, 2018.

[2] C. P. L. van Vroonhoven, and K. A. A. Makinwa, "A CMOS Temperature-to-Digital Converter with an Inaccuracy of ± 0.5° C (3/spl sigma)from -55 to 125°C," in *IEEE International Solid-State Circuits Conference*, San Francisco, 2008.

[3] R. Quan, U. Sonmez, F. Sebastiano, and K. A. A. Makinwa, "A 4600μm² 1.5°C (3σ) 0.9kS/s thermal-diffusivity temperature sensor with VCO-based readout," in *IEEE International Solid-State Circuits Conference*, San Fransisco, 2015.

[4] M. Shahmohammadi, K. Souri, and K. A. A. Makinwa, "A resistor-based temperature sensor for MEMS frequency references," in *ESSCIRC*, Bucharest, 2013.

[5] U. Sönmez, F. Sebastiano, and K. A. A. Makinwa, "Analysis and Design of VCO-Based Phase-Domain ΣΔ Modulators," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 5, pp. 1075-1084, 2017.

[6] J. Angevare, L. Pedalà, U. Sönmez, F. Sebastiano, and K. A. A. Makinwa, "A 2800-μm² thermal-diffusivity temperature sensor with VCO-based readout in 160-nm CMOS," in *IEEE Asian Solid-State Circuits Conference*, Xiamen, 2015.

[7] U. Sönmez, F. Sebastiano, and K. A. A. Makinwa, "11.4 1650μm² thermal-diffusivity sensors with inaccuracies down to ±0.75°C in 40nm CMOS," in *IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, 2016.

[8] W. Choi, Y. Lee, S. Kim, S. Lee, J. Jang, J. Chun, K. A. A. Makinwa, and Y. Chae, "A Compact Resistor-Based CMOS Temperature Sensor With an Inaccuracy of 0.12°C (3σ) and a Resolution FoM of 0.43 pJ·K² in 65-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 12, pp. 3356-3367, 2018.

[9] Y.-C. Hsu, C.-L. Tai, M.-C. Chuang, A. Roth, and E. Soenen, "5.9 An 18.75μW dynamic-distributing-bias temperature sensor with 0.87°C (3σ) untrimmed inaccuracy and 0.00946mm² area," in *IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, 2017.

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

[11] J. A. Angevare and K. A. A. Makinwa, "A 6800-μm² Resistor-Based Temperature Sensor With ±0.35 °C (3σ) Inaccuracy in 180-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 10, pp. 2649-2657, 2019.

Chapter 4 | Design II: Digital Readout & Limit Cycling

This chapter describes an improved version of the RC-based temperature sensor described in the previous chapter [1]. This included a CCO in the signal path between its Wien Bridge (WB) filter and its phase ADC (Figure 4.1). As a result, its accuracy is limited by CCO non-linearity. To avoid loading the WB filter, the CCO was driven via a current buffer, which adds extra analog complexity and limits scalability. To address these issues, the main goals of this design are 1) to reduce the reliance on analog circuitry, and 2) to remove the CCO from the signal path and thus improve sensor accuracy.

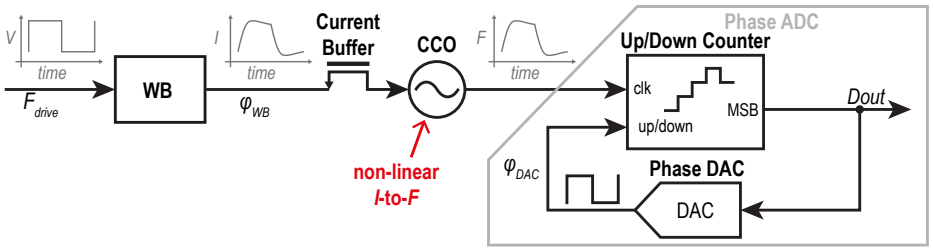

Figure 4.1: Simplified system diagram of previous design, showing a non-linear CCO in the signal path.

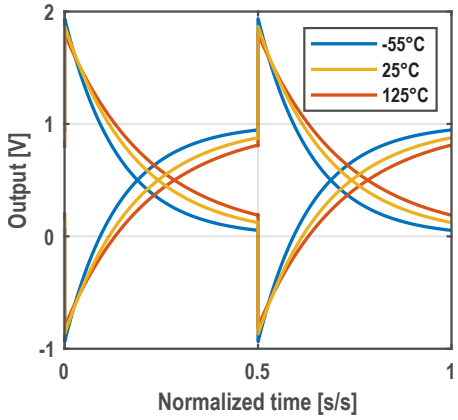

Figure 4.2: PPF transient response for different temperatures and V_{DD} = 1 V.

In the previous design, the CCO was used to convert the analog output of the WB filter into the quasi-digital signal required by the counter/integrator of a highly-digital PDΣΔM. Alternatively, this can be done by using a comparator to detect the zero-crossings (ZCs) in the filter's output (see Section 2.2.2.3). The CCO can then be removed from the signal path. In this case, an RC filter with a large output and strongly temperature-dependent ZCs should be used (Figure 4.2), leading to the choice of a poly-phase filter (PPF) rather than a WB filter (Ch. 2.2.2.2.). The comparator then becomes the most critical part of the design, since it must accurately detect the ZCs over PVT (see Section 4.4).

Figure 4.3 shows the proposed system diagram. In it, a poly-phase filter (PPF) is driven by a square wave (F_{drive}) resulting in ZCs that are detected by a comparator. The analog phase information in the ZCs is then digitized by a PDΣΔM. An XOR gate acts as the phase summation node of the PDΣΔM, which is possible because both its inputs are square waves: the

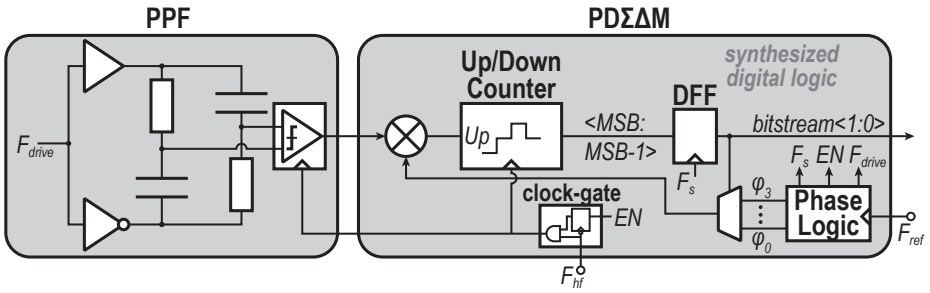

Figure 4.3: Simplified system diagram.

output of the comparator and the reference square-wave selected by the modulator. The output of the XOR gate is a square-wave whose duty cycle is linearly related to the phase difference between its inputs. A counter clocked at a fixed (high) frequency integrates the square-wave and, via a D flip-flop (DFF) that acts as a comparator, drives the average phase of the reference square wave to be 90° out of phase with respect to the comparator's output. A phase logic block generates all required phase signals from a reference clock (F_{ref}).

4.1: Digital Mixer

The main advantage of the proposed architecture is that the CCO is no longer required in the signal path, and the output of the phase detector can be directly integrated via the counter's up/down port. The counter's clock frequency is now fixed and can be generated by a simple oscillator that is shared between multiple temperature sensors.

A second advantage is that the XOR mixer has a linear phase characteristic. Previous PDΣΔMs used choppers to synchronously detect the phase of near-sinusoidal signals, and so suffered from the so-called cosine non-linearity [2]. The presence of this non-linearity complicates the calibration and trimming of errors due to resistor spread, and also introduces significant noise-folding in the case of multi-bit PDΣΔMs (Ch. 2.3.2.1). In contrast, the duty-cycle of the XOR output is linearly dependent on the input phase. The value that is accumulated in each ΣΔM-period (ΔC) is then given by:

$$\Delta C = \frac{2}{180} \cdot \frac{F_{CCO}}{F_{drive}} \cdot \Delta\varphi \qquad (4.1),$$

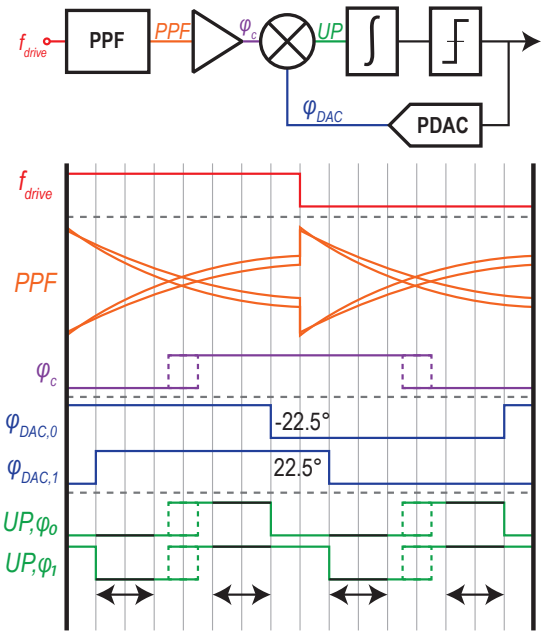

Figure 4.4: Operation of a 1-bit PDΣΔM with PPF under extreme conditions, showing signal independent segments in the PDΣΔM's error signal (*UP*).

where $\Delta\varphi$ is the input phase shift minus 90°, F_{CCO} is the CCO frequency, and F_{drive} is the PPF driving frequency.

4.2: Windowing

In this design, both the comparator and the counter are the primary power consumers. However, they can be strategically disabled (windowed) to save power. Figure 4.4 shows the operation and ZCs of a 1-bit PDΣΔM at two temperature extremes. Large periods of time can be identified (black arrows) where the state of the counter's up/down signal (*UP*) is the same for both DAC phases ($\varphi_{DAC,x}$). This means that these periods do not contain any signal-dependent information. Consequently, both the comparator and the counter can be windowed during these periods.

Windowing is equivalent to demodulation with a three-level waveform, whose middle level is zero and results in no integration [3]. This three-level waveform is a closer approximation of a sine wave. However, the linearity of the XOR mixer (Section 4.2) is maintained as long as the counter is only windowed during signal-independent time segments.

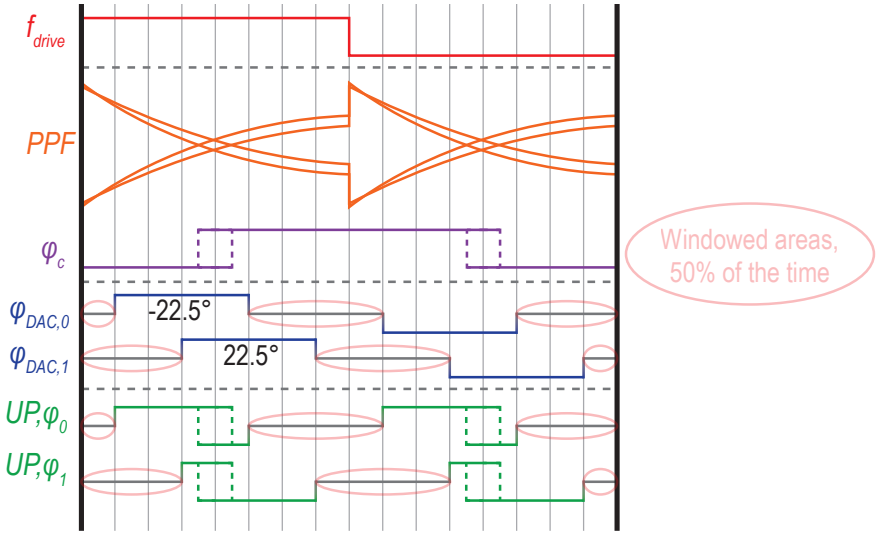

Figure 4.5: Straightforward windowing implementation with 50% power saving.

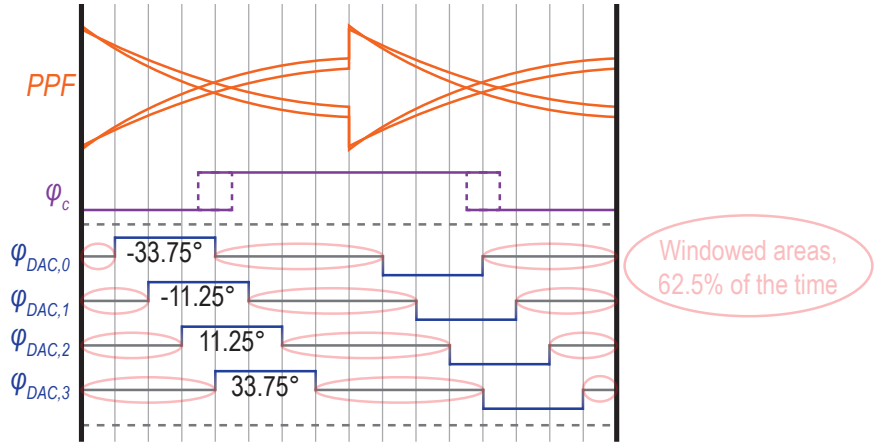

Figure 4.6: Optimized multi-bit windowing with 62.5% power saving.

A straightforward windowing implementation involves choosing the two phase-DAC references such that they straddle the ZC phase range over temperature (Figure 4.5). However, this only results in a 50% power saving. Alternatively, a multi-bit ΣΔM can be used, which narrows the ZC phase range that is seen by the separate feedback phases ($\varphi_{DAC,x}$). As a result, the zero-level duration and the amount of power saved can be increased to 62.5% (Figure 4.6). The non-zero levels of the consecutive feedback phases still overlap, so the modulator can always toggle between input-sensitive feedback phases.

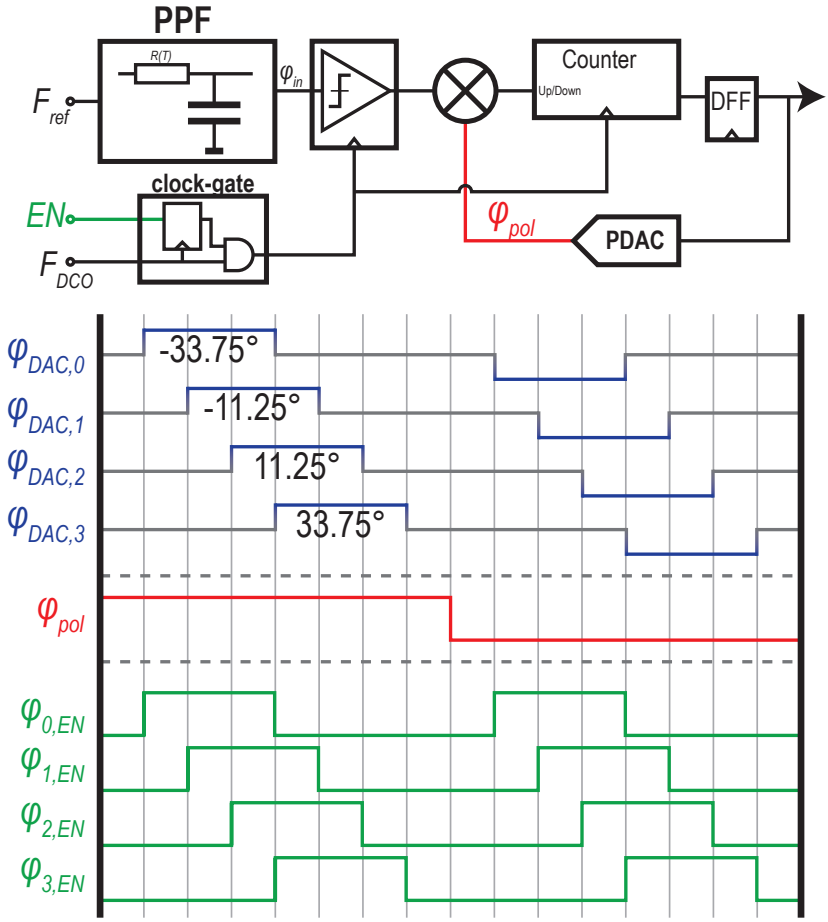

Figure 4.7: Three-level waveform composition.

Figure 4.7 shows the realization of the three-level demodulation scheme with the help of two digital signals. An enable signal (*EN*) defines the transitions, while a polarity signal (φ_{pol}) defines their polarity. Since the enable signal (*EN*) is bitstream-dependent, a clock gate is used to sysnchronize it to the CCO's frequency, thus avoiding meta-stability in the multi-stage counter.

4.2.1 – Time Quantization Noise

The square-wave demodulation (Section 4.2) and the windowing scheme alter the noise originating from the counter's time quantization (Ch. 2.3.2.4). Traditional VCO-based PDΣΔMs have two up/down transitions

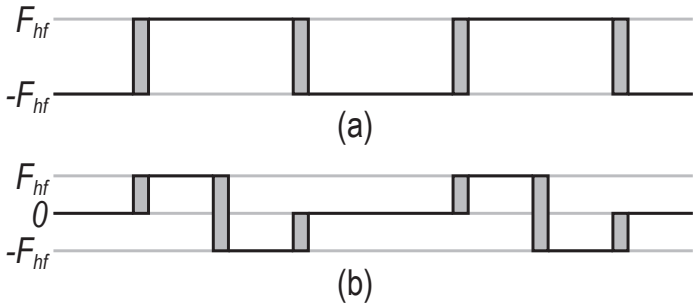

Figure 4.8: Time-discretization noise in (a) square wave demodulation and (b) three-level demodulation.

every ΣΔM period [4], while two-level demodulation has four (Figure 4.8 (a)). The windowing scheme changes this to six transitions, but these transitions are smaller in amplitude (Figure 4.8 (b)). For the traditional VCO-based PDΣΔM, we obtained (Ch. 2.3.2.4, Eq. (2.27)):

$$\sigma_{\Delta C}{}^2 = 2 \cdot \left(\frac{1}{\sqrt{12}} \cdot T_{CCO} \cdot 2 \cdot f_{CCO}\right)^2 = \frac{2}{3} \quad (4.2),$$

where $\sigma_{\Delta C}$ is the variation in counts that are integrated in the counter each ΣΔM period, and T_{CCO} and f_{CCO} are the CCO period and frequency. When we modify this for the two-level case, we obtain:

$$\sigma_{\Delta C}{}^2 = 4 \cdot \left(\frac{1}{\sqrt{12}} \cdot T_{hf} \cdot 2 \cdot f_{hf}\right)^2 = \frac{4}{3} \quad (4.3),$$

where T_{hf} and f_{hf} are the counter clock period and frequency. Finally, for the windowed case we obtain:

$$\sigma_{\Delta C}{}^2 = 2 \cdot \left(\frac{1}{\sqrt{12}} \cdot T_{hf} \cdot 2 \cdot f_{hf}\right)^2 + 4 \cdot \left(\frac{1}{\sqrt{12}} \cdot T_{hf} \cdot f_{hf}\right)^2 = 1 \quad (4.4).$$

Clearly, the windowed case causes less noise than the two-level case as a result of the decreased correlation between the transition errors. Its noise is still higher than the traditional case, but its sensitivity to the noise is also lower since its peak-to-peak CCO frequency swing is smaller than that of the two-level f_{hf} (Eq. (2.31), Table 2.2).

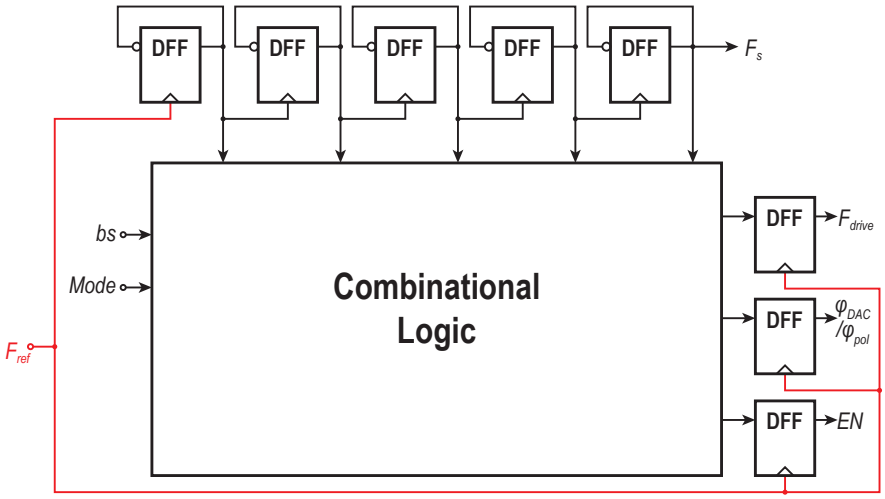

Figure 4.9: Simplified schematic of the phase logic.

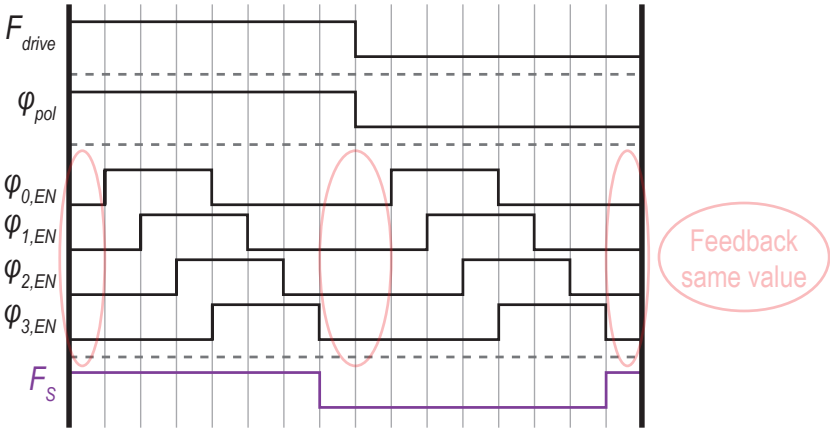

Figure 4.10: Timing diagram of the phase logic.

4.3: Phase Logic

Figure 4.9 shows the phase logic, which uses a higher reference frequency (F_{ref}) to generate the drive signal (F_{drive}), the feedback phases (φ_{DAC} or φ_{pol}), the enable/disable signal (EN) of the three-level waveform, and the sampling signal (F_s). It consists entirely of synchronous, synthesized digital logic. A ripple counter-like structure down-converts F_{ref} to the desired F_s, as shown at the top of Figure 4.9. To evaluate the effect of windowing, the phase logic can be set to different modes (un-windowed, windowed-with-zoom,

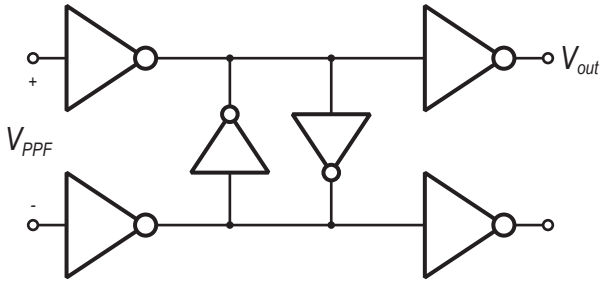

Figure 4.11: Circuit diagram of the static comparator used in [5,6].

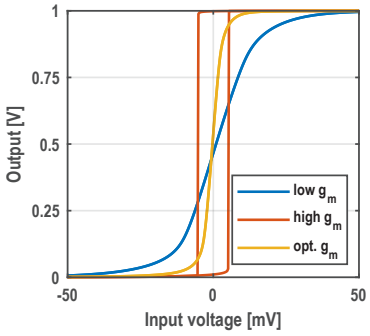

Figure 4.12: Comparator output for different relative latch gm values.

and multi-bit windowed). A combinational block then uses the frequency divider states, the bitstream value, and the phase logic mode to generate the phase signals. To mitigate timing errors due to delays in this block, all phase signals are reclocked to F_{ref}. The sampling moment (F_s) is chosen such that a toggling bitstream does not result in a toggling feedback signal, i.e., the sampling moment occurs when φ_1, φ_2, φ_3, and φ_4 all have the same level, see Figure 4.10.

4.4: Dynamic Comparator

The prior art uses a static comparator that consists of two inverters driving a cross-coupled latch [5,6] (Figure 4.11). If the latch transconductance is not much smaller than that of the driving inverters, the comparator will exhibit hysteresis (Figure 4.12), which shows up as a PPF slope-dependent delay:

$$P_{e,hys} = \frac{360^\circ}{T_s} \cdot \tau \cdot \ln \frac{V_{DD}}{V_{DD} - 2 \cdot V_{hys}} \quad (4.5),$$

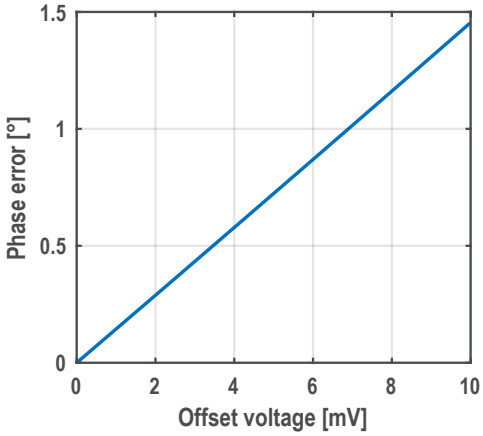

Figure 4.13: Phase error due to comparator hysteresis, for τ = 0.1 µs, Ts = 0.5 µs, and V_{DD} = 1 V.

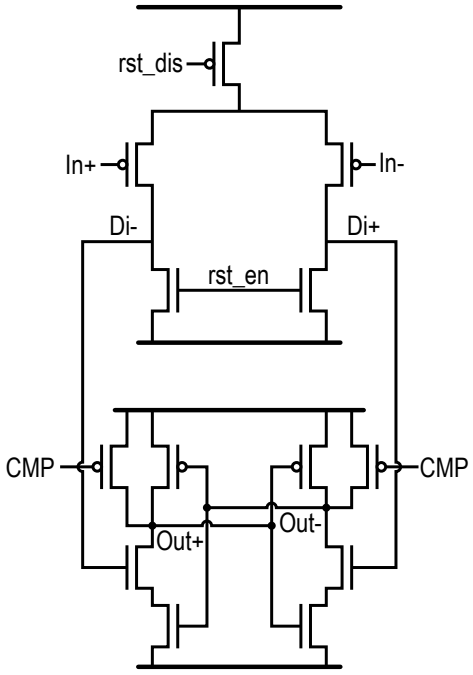

Figure 4.14: Simplified schematic of a double-tail comparator [7] with Elzakker latch [9].

where $P_{e,hys}$ is the phase error due to hysteresis, T_s is the modulator period, τ is the PPF time constant, V_{DD} is the supply voltage, and V_{hys} is the hysteresis voltage. Figure 4.13 shows that hysteresis has a severe effect

on the inaccuracy of the sensor. Therefore, the latch transconductance is designed to be much smaller than that of the driving inverters. However, if it is too small, the system will behave like a single-ended system, resulting in a loss of accuracy and PSS. The optimum point is where the latch is just short of hysteresis. However, as PVT varies, the hysteresis changes, causing increased inaccuracy and PSS in the latch.

In contrast, a dynamic comparator is completely differential, resulting in much better accuracy and PSS. Although it consumes more power than a static comparator, the total power consumption will still be dominated by the counter's power consumption. One downside is that it only samples the PPF output at discrete moments in time. However, since the counter/integrator does the same, this does not increase the time quantization noise. Furthermore, it exhibits much less hysteresis, and as a result can achieve better accuracy.

Since the dynamic comparator should be low-power, low-offset, and decide fast, a double-tail comparator [7] with an Elzakker latch [8] is used (Figure 4.14). During the reset, the internal nodes Di+ and Di- are reset to ground. When the compare signal is given, the input transistors charge these nodes to V_{DD}. Due to a difference in input voltage and the input transistors' transconductance, one side charges faster. Once either of the internal nodes charges past the threshold voltage of the latch's input transistors, the latch starts latching. This helps reduce the power consumption and the comparator offset, since the internal node has time to integrate the input voltage before the latch starts latching.

4.4.1 – Offset

To first order, the timing error due to comparator offset made during succesive rising and falling ZCs will cancel out [6]. However, the (non-linear) exponential settling of the PPF causes a non-symmetric cancellation, and so there is some residual offset:

$$P_{e,os} = \frac{360^\circ}{T_s} \cdot \tau \cdot \ln \frac{V_{DD}}{\sqrt{{V_{DD}}^2 - 4 \cdot {V_{os}}^2}} \tag{4.6},$$

where $P_{e,os}$ is the phase error due to comparator offset, T_s is the modulator period, τ is the PPF time constant, V_{DD} is the supply voltage, and V_{os} is the offset voltage. Unsurprisingly, the polarity of the offset voltage does

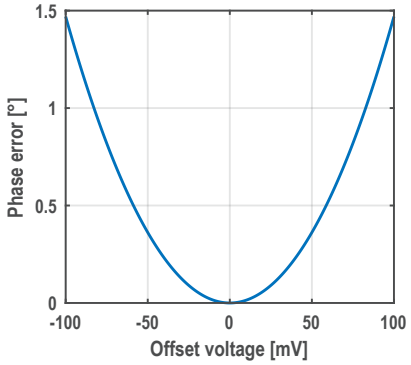

Figure 4.15: Phase error due to comparator offset, for $\tau = 0.1\,\mu s$, $Ts = 0.5\,\mu s$, and $V_{DD} = 1\,V$.

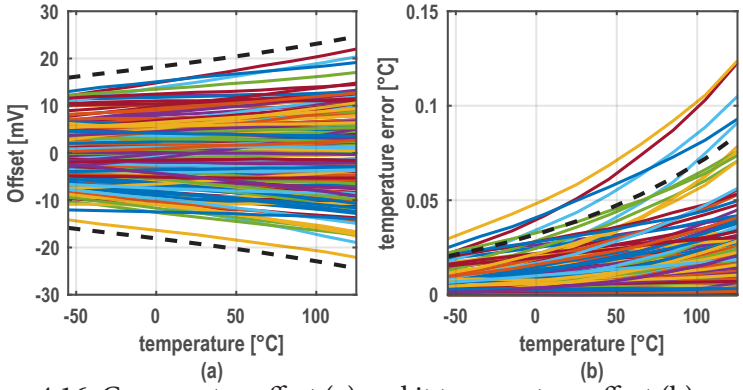

Figure 4.16: Comparator offset (a) and it temperature effect (b).

not matter. Also, the effect of comparator offset (Figure 4.15) is much smaller than that of comparator hysteresis (Figure 4.13). Figure 4.16 shows the simulated offset of the comparator and its effect on the measured temperature (obtained from a Monte-Carlo simulation). The black dashed lines indicate the 3σ error, which shows that offset has a negligible effect on inaccuracy (<0.1°C).

4.4.2 – Delay

As shown in Figure 4.17, errors in both the signal and demodulation paths contribute to errors in the measured PPF phase (delay). The drive path is implemented by the phase logic and contains: a clock-to-q delay, a buffer delay, the PPF delay, and the data-to-clock delay of the comparator. In the case of windowed three-level demodulation (see Section 4.2), this delay is balanced by the enable delay, which contains: a clock-to-q delay, an EN-to-clock delay of the clock gate, and an AND-gate delay. To first-order, the

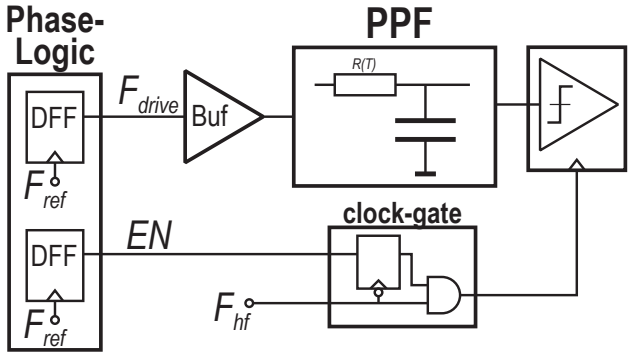

Figure 4.17: Delay paths that contribute to the measured phase.

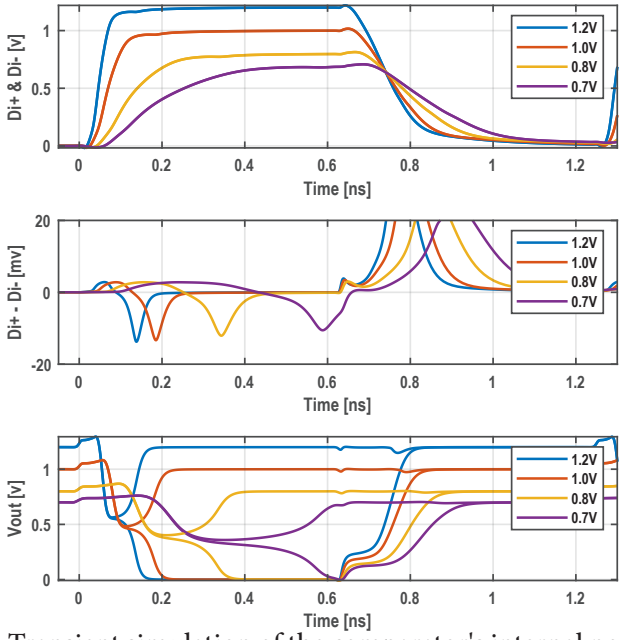

Figure 4.18: Transient simulation of the comparator's internal node (top), the comparator's internal node differential voltage (middle), and the comparator's output node, which show different input-to-clock delays for different supply voltages.

clock-to-q delays, buffer, and AND-gate delay cancel each other. However, the comparator's data-to-clock delay is considerable compared to the clock gate's EN-to-clock delay. This is because the comparator requires time to integrate enough differential charge on its internal nodes Di+ and Di- (Figure 4.18 (middle)) and for the latch to change state (Figure 4.18 (bottom)). This delay is sensitive to PVT, so it should be reduced to minimize its effect on the sensor's accuracy and PSS. In this design, the comparator's phase

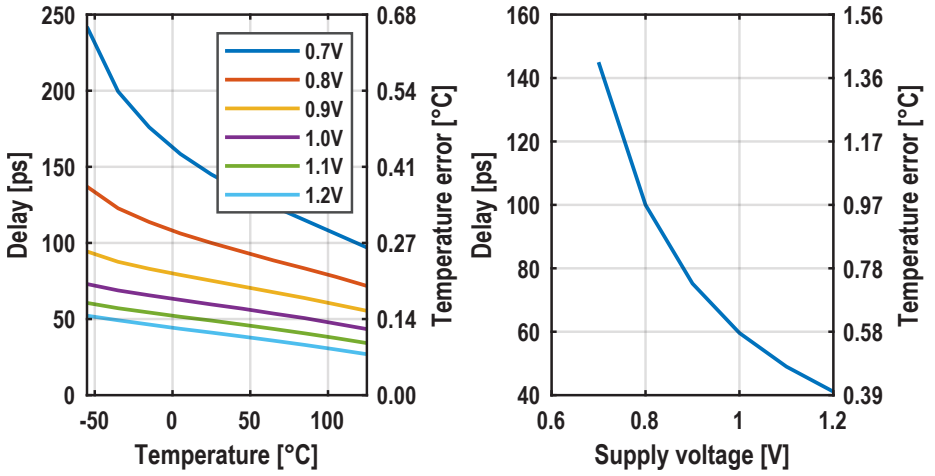

Figure 4.19: Simulated comparator delay and equivalent temperature error over temperature and supply.

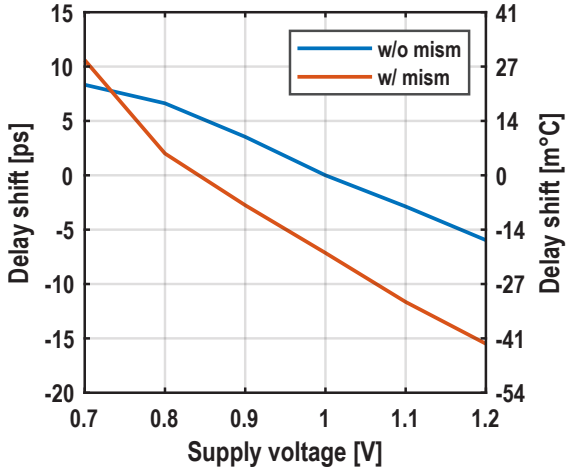

Figure 4.20: Effect of charge kickback in the dynamic comparator.

delay is responsible for a temperature error of <0.5°C and a PSS of 0.6°C/V (Figure 4.19). Reducing this delay results in a larger comparator offset and noise sensivity, since it reduces the time over which the differential voltage in the internal node is integrated. It should be noted that a significant proportion of the delays are caused by digital logic which will decrease in scaled technologies.

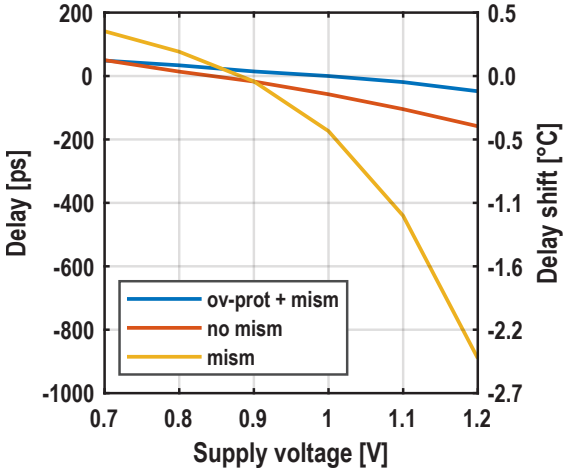

Figure 4.21: Simulated effect of comparator gate leakage on the measured delay.

4.4.3 – Charge Kickback

To simplify the design, the dynamic comparator does not use a pre-amp. This means that the voltage swing in the internal node (Di+ and Di-) causes charge kickback in the comparator input (in+ and in-), i.e., in the PPF. To first order, Di+ and Di- transition by the same amount, meaning there is no differential effect. However, offset in the comparator input transistors, and the kickback difference due to the differential voltage on In+ and In-, cause a residual effect (Figure 4.20). This effect is relatively small (<50 mK) and is insignificant compared to other sources of error.

4.4.4 – Overvoltage

The PPF swing is almost three times as large as the supply voltage (V_{DD}). Consequently, the input transistors of the comparator have to handle voltages as large as $2 \cdot V_{DD}$ across their gate oxides. This results in a reduced transistor lifetime and an exponentially larger gate leakage/tunneling current. This current adds to the PPF current and thus degrades sensor performance. This is illustrated in Figure 4.21, which shows the supply sensitivity due to the comparator's gate leakage. With mismatch, its sensitivity is more than 5.4°C/V. To reduce their gate leakage and increase their lifetime, the comparator's input transistors are temporarily shorted to

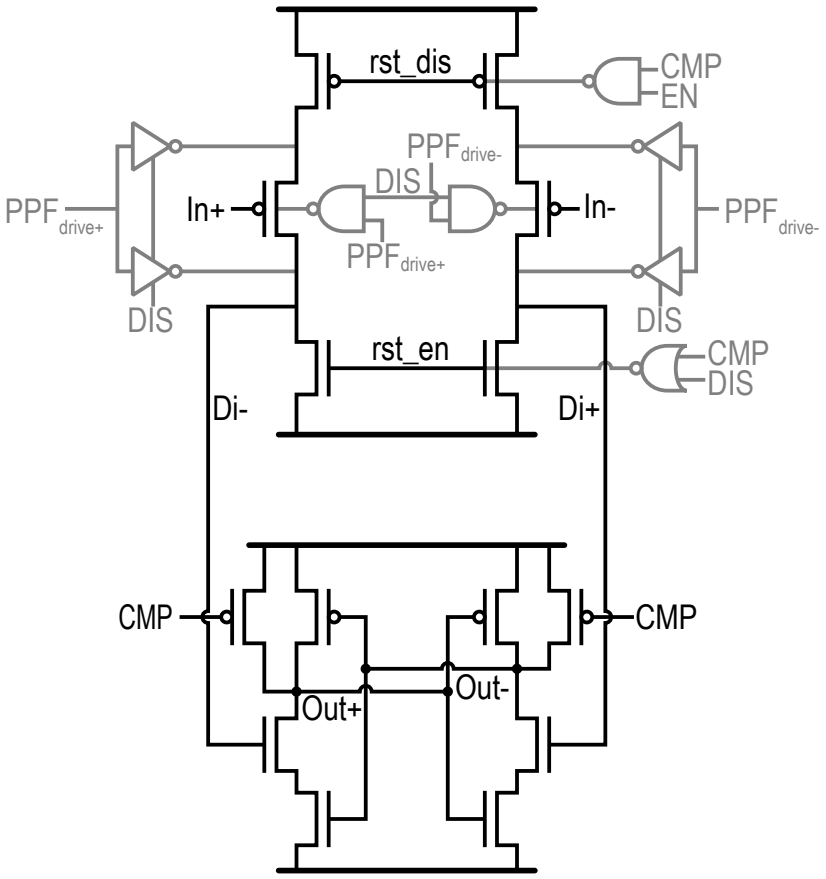

Figure 4.22: Schematic of the double-tail comparator with overvoltage protection and an Elzakker latch.

V_{DD} or ground during the overvoltage moments, thus ensuring that their gate oxide voltage is always less than V_{DD}. This technique can be readily combined with the comparator's windowing (see Section 4.2), which conveniently disables the comparator during the PPF's overvoltage. The extra circuitry required to short the input transistors is shown in Figure 4.22. Using this overvoltage protection scheme, the gate leakage effect is reduced to 0.5°C/V (Figure 4.21).

Finally, Table 4.1 gives an overview of the contributions of the error sources of the dynamic and static comparator.

Table 4.1: Error contributions of the dynamic and static comparator.

	Dynamic Comparator		Static Comparator	
Contributor	Error [°C]	Supply Sensitivity [°C/V]	Error [°C]	Supply Sensitivity [°C/V]
Hysteresis	—	—	0.003	0.02
Offset	0.1	—	0.3	0.11
Delay	0.5	0.6	3.0	7.5
Charge Kickback	0.05	0.14	—	—
Overvoltage	—	0.5	—	—

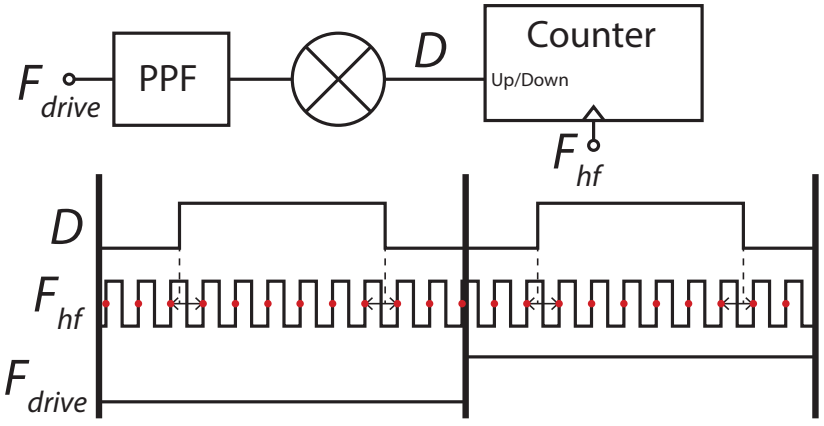

Figure 4.23: Illustration of the counter's operation.

4.5: Beat Period

The counter detects the XOR gate's duty cycle (D). For this, it has to detect the moment of D's rising and falling edge. However, it can only detect D's level at the rising edges of its high frequency counter clock F_{hf}, see Figure 4.23. After one period, it only knows that the rising edge happens between the third and fourth clocking moment while the falling edge happens between the ninth and tenth clocking moment. As the conversion continues, the counter gets more information and the timing uncertainty that is due to the time quantization decreases.

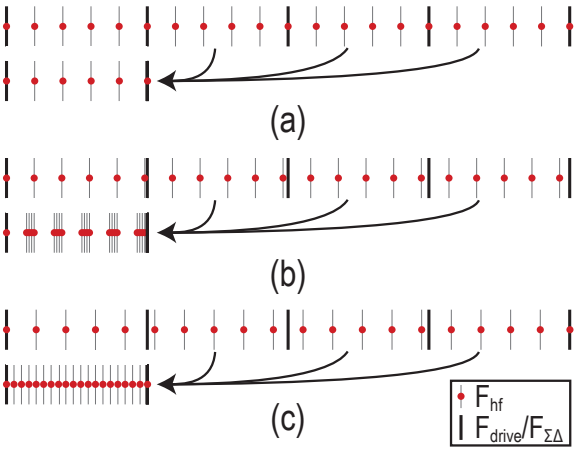

Figure 4.24: Effect of the re-alignment interval between F_{hf} and F_{drive}, where (a) re-aligns after one period of F_{drive}, (b) never re-aligns, and (c) re-aligns after a complete conversion.

The ratio between F_{drive} and the counter's clock frequency (F_{hf}) greatly influences the counter's time-domain quantization noise. A near-integer ratio causes correlated, rather than stochastic, errors, which in turn causes severe temperature sensing errors. This is illustrated in Figure 4.24 (a), where F_{drive} and F_{hf} are chosen such that they are aligned at the start of each F_{drive} period. After aligning all the ΣΔM periods of one conversion, we can see that all clocking moments (F_{hf}) align. As a result, differences in the ZC phase between those moments cannot be detected, resulting in a time-domain induced dead zone. Choosing an irrational ratio between F_{drive} and F_{hf} also does not guarantee good phase sensitivity (Figure 4.24 (b)). In this case, the transitions of F_{drive} and F_{hf} will only re-align after a very long period, which can be much longer than the intended conversion time. Still, if we align all ΣΔM periods, it is clear that there are still large gaps of zero phase sensitivity, i.e., time-domain induced dead zones.

To better understand this phenomenon, the system can be analyzed in terms of its so-called beat frequency. This is the period required for both tones in a two-tone system to re-align. This is an important property in any two-tone system, e.g., in music where a short beat period means a pleasant interval, in a two-tone linearity test where the simulation time should be a multiple of the beat period to prevent spectral leakage, and in this system where an optimal beat period guarantees good uniform ZC phase sensitivity. To minimize the time-domain induced dead zones, we should maximize the beat period, although the beat period should not be

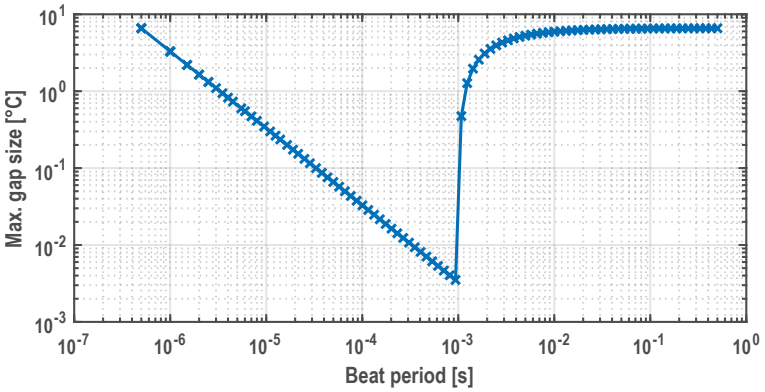

Figure 4.25: Time-domain induced dead zone size for different beat frequencies of F_{hf} and F_{drive}, with T_{conv} = 1 ms, F_{drive} = 2 MHz, and $F_{hf} \approx$ 590 MHz.

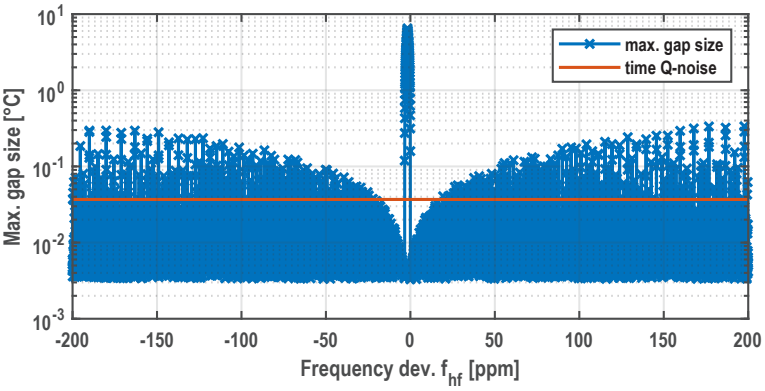

Figure 4.26: F_{hv}/F_{drive} frequency ratio-induced dead zone size and time-domain Q-noise for small deviations of F_{hf} around an optimal point ($F_{hf} \approx$ 590 MHz, F_{drive} = 2 MHz, and T_{conv} = 1 ms).

longer than the conversion time to prevent the situation shown in Figure 4.24 (b). This means that to minimize the dead zones, we should maximize the F_{hf}/F_{drive} ratio and set their beat period equal to the conversion time. This is confirmed in Figure 4.25, which shows that the dead zones decrease in size as the beat period approaches the conversion time. Once the beat period increases past the conversion time, the width of the dead zones rapidly increase up to a value determined by $1/F_{hf}$.

The main problem with these time-domain induced dead zones is illustrated in Figure 4.26. In this figure, F_{hf} is tuned to an optimal frequency (i.e., where the beat frequency is equal to the conversion time) and then swept with tiny increments over a ±200 ppm range. It shows a main lobe where the dead zone is very large, and, outside of it, a complex pattern with smaller

Table 4.2: Different sensor versions.

Version	Resistor Width	Comparator
1	400nm	Dynamic
2	400nm	Static
3	200nm	Dynamic
4	200nm	Static

and larger dead zones which become increasingly larger. The main lobe is located directly to the left of the optimal beat period and corresponds to a near-perfect rational ratio between F_{hf} and F_{drive}, i.e., 590 MHz / 2 MHz. The complex pattern outside the main lobe results from the intricate beat period changes as the phases of F_{hf} sweep along the phases of F_{drive}. The red line indicates the resolution limit due to the time-domain Q noise. As long as the dead zones are smaller than the resolution, they are not measurable.

To guarantee this, an oscillator with a frequency deviation better than 20 ppm over PVT is required. Noise in the oscillator helps to mask this to some degree. However, drift and temperature variations will invariably result in an inappropriate frequency ratio. Furthermore, injection locking will exacerbate the beat period problem, as it will lock F_{hf} to the harmonics of F_{drive}. The effect of this will be worse in the dense and high-power circuitry that requires thermal management. In practice, when using a free-running oscillator, dead-zones that are as large as ~10°C have been measured. At some temperatures, this will be more likely to happen.

The solution is to employ a fractional PLL that tightly locks F_{hf} to F_{drive}, i.e., it forces $F_{hf} = M/N \cdot F_{drive}$, where M and N are integer numbers. It can then ensure that F_{hf} re-aligns with F_{drive} at the end of the conversion by setting N equal to the decimation length (2048 in our case). F_{hf} can then be set as high as required by setting M. However, to ensure that the F_{hf}/F_{drive} cycle does not repeat more than once during a conversion, M and N must not have a common divisor. In this work (unless noted otherwise) we use F_{drive} = 1.875 MHz (which centers the PPF phase-shift), M = 622085, N = 2048, and so F_{hf} = 569,535,827.6 Hz. This is done using a Silicon Labs PLL [9].

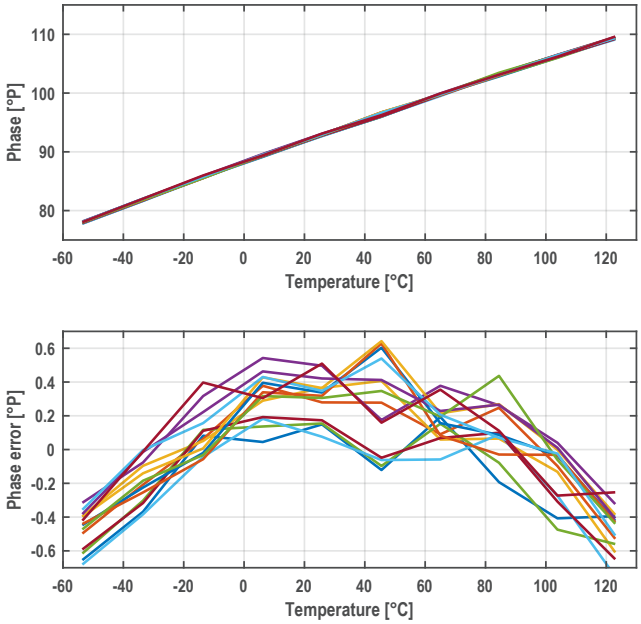

Figure 4.27: Measured phase (top) and phase error after a systematic first-order fit (bottom).

4.6: Measurement Results

Four different versions of the PPF-based temperature sensor were taped-out. To measure the improvement of the dynamic over the static comparator in [5,6], both were implemented. The PPF in [5,6] employs a larger than minimum width resistor, but larger PPF time constants lead to reduced phase delay effects from the readout circuitry and a more accurate design. As a result, two resistor versions were taped-out, one with the same PPF time constant (37 ns) and the same resistor width as [6] (400 nm), and, as an improvement, another version with a minimum width resistor (200 nm) using the same area and hence a larger time constant (113 ns). Table 4.2 sumarizes the different sensor versions. Unless otherwise noted, measurement results are from version 3 (200 nm-wide resistor & dynamic comparator).

4.6.1 – Ringing

Figure 4.27 shows the measured phase over temperature with steps of 20°C for 14 samples of version 3 from -55°C to 125°C. After a first-order fit, large jumps in the measured phase can be seen. The cause of these jumps was

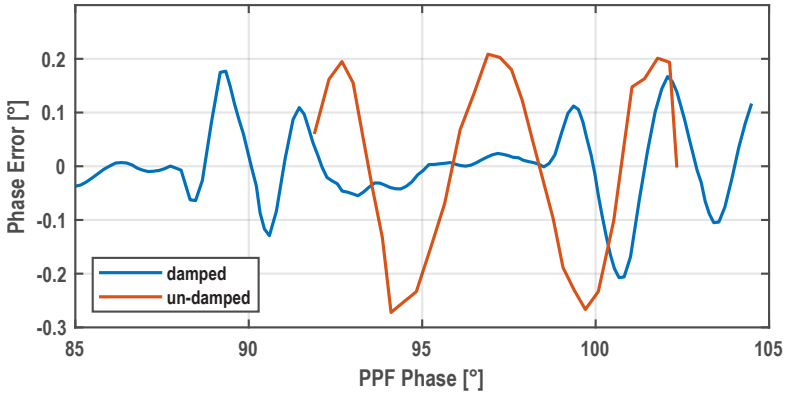

Figure 4.28: Oscillations in the phase readout.

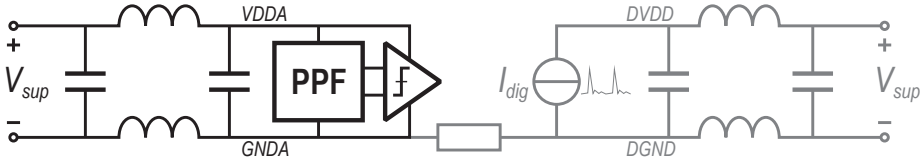

Figure 4.29: Parasitics that cause ringing in the power supply (black shows the ringing in the analog power domain, whereas gray shows the ringing in the digital power domain and how it couples to the analog domain).

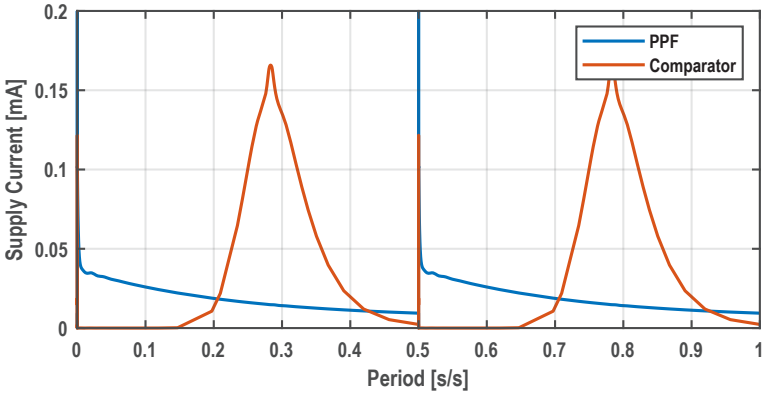

Figure 4.30: Simulated supply current in the analog power supply.

traced to a systematic ripple in the sensor's output, as shown in Figure 4.28, which shows the measured phase deviation from a first-order fit over a narrow phase range. This ripple was traced to ringing in the sensor's analog power supply, which mainly consist of the bond-wire inductances and an on-chip bypass capacitor (Figure 4.29, black circuit only). Most of the

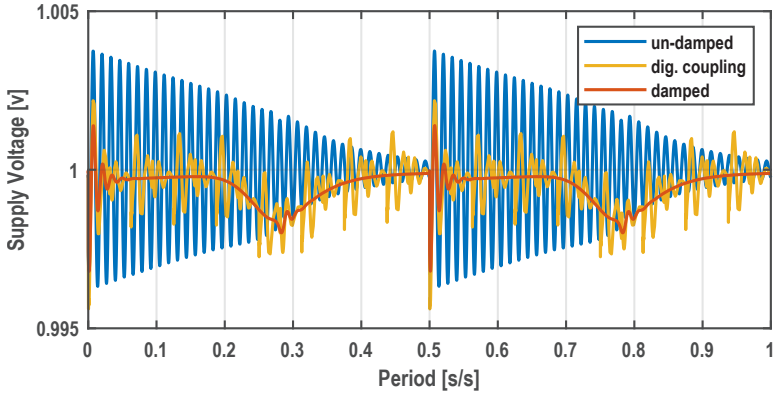

Figure 4.31: Simulated power-supply ringing due to bond-wire parasitics.

ringing is triggered by periodic spikes in the sensor's analog supply current, which occur when the PPF's input toggles and when the comparator toggles during a ZC (Figure 4.30). The resulting (simulated) ringing of the analog power supply (V_{DD}) is shown in Figure 4.31. The ringing can be reduced by inserting a 10 Ω resistance in series with V_{DD}, which ensures that the ringing due to PPF toggling is damped before the resulting ZC (Figure 4.31 (red)).

The chips were re-measured with this 10 Ω series resistance. Figure 4.28 shows the measured phase deviation from a 2nd order fit (in blue) which shows damped ringing around 90° and 112.5°. This is caused by ringing in the digital supply, which couples into the analog domain since the sensor is quite small and its analog and digital domains are very close together. As a result, the substrate resistance between them is quite low (modeled by the gray circuitry in Figure 4.29). The digital supply experiences spikes at F_{ref} (which is 16x F_{drive}) and at the CCO's frequency. The latter is fairly random with respect to F_{drive}, so these spikes effectively increase the TQ noise, but do not have a systematic effect. In contrast, the ringing due to the spikes at F_{ref} do result in a repetitive pattern, which is shown in Figure 4.31 (yellow). This ringing pattern affects the output of the PPF, which acts as an all-pass filter for signals coming from the power supply. These transient signals can thus directly affect the PPF output voltage and consequently change the ZC moment. This is a significant issue, since the sensor is intended to be embedded in dense digital circuitry, which will inevitably lead to ringing in its supply voltage. Using Eqs. (2.12) and

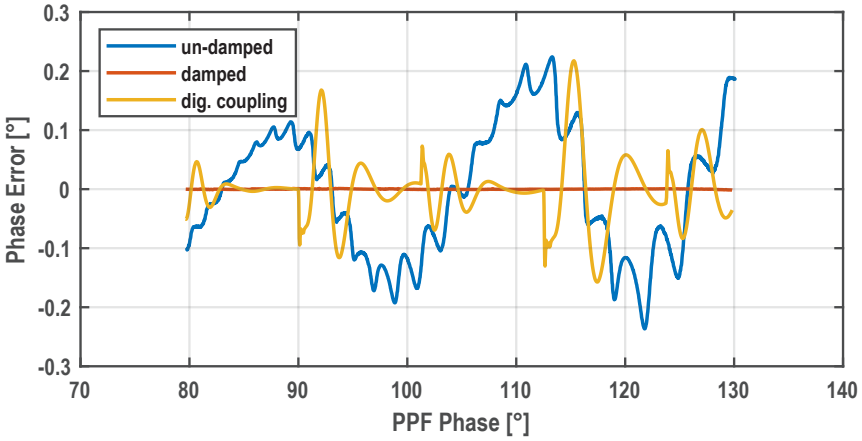

Figure 4.32: Simulated phase error caused by ringing.

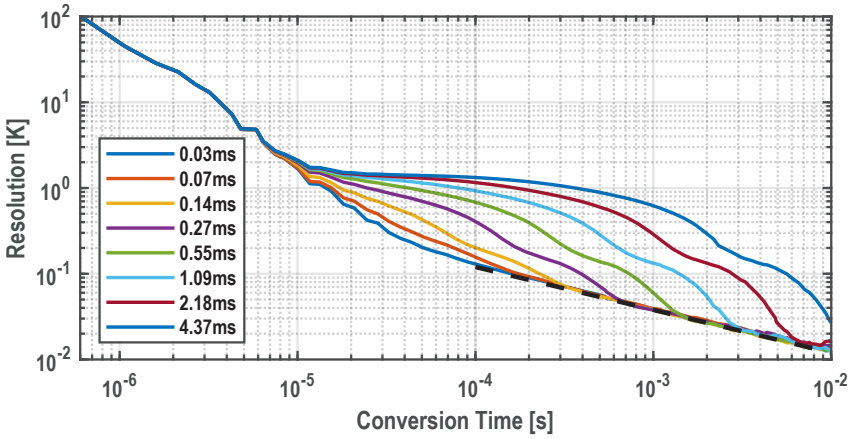

Figure 4.33: Resolution vs. conversion time for different beat periods.

(2.14), we find that even small oscillations in the power supply (2 mV) will cause significant oscillations in the measured temperature (0.4°, which corresponds to 2.1°C). Figure 4.32 shows the simulated phase error due to this ringing under the same conditions as presented in Figure 4.31. Both the simulations and the calculation correspond well to the damped system's measured phase error (Figure 4.28 (blue)). This error in the phase measurement cannot be removed by trimming or fitting, which means that this phenomenon limits the sensor's accuracy.

Table 4.3: Frequencies used in Figure 4.33 where *M* and *N* are the amount of F_{hf} and F_{drive} periods within one beat period respectively.

F_{hf} [MHz]	*M*	F_{drive} [MHz]	*N*	Beat Period [ms]
555.205,078,1	18,951	1.875	64	0.03
555.073,242,2	37,893	1.875	128	0.07
555.036,621,1	75,781	1.875	256	0.14
555.018,310,5	151,557	1.875	512	0.27
555.009,155,3	303,109	1.875	1,024	0.55
555.004,577,6	606,213	1.875	2,048	1.09
555.002,288,8	1,212,421	1.875	4,096	2.18
555.001,144,4	2,424,837	1.875	8,192	4.37

As a result of the ringing in the digital supply and its coupling into the analog domain, the intrinsic accuracy of this sensor could not be measured. This is shown in Figure 4.28 (blue), which shows peak-to-peak oscillations that correspond to a temperature error of 2.1°C. Interestingly, other publications that use the ZC of a PPF for a temperature sensor do not seem to suffer from this issue [5,6]. There are a three possible explanations: 1) they do not multiplex as many temperature sensors on one chip which leaves more space for bypass caps, 2) they use more ground pins, and 3) they use smaller packages which have smaller bond-wires/leads which exhibhit smaller inductance.

4.6.2 – Resolution

Figure 4.33 shows the resolution vs. conversion time for different beat periods between F_{drive} and F_{hf} where the dashed line is the theoretical noise floor. As explained in Section 4.5, the beat period is the period of time after which the time quantization noise pattern repeats. This means that the sampling window size that can effectively filter out this pattern changes with the beat period, which is confirmed by Figure 4.33. The frequencies used to generate this plot are shown in Table 4.3. The M and N columns indicate the amount of periods of F_{hf} and F_{drive} that fit within the beat period respectively.

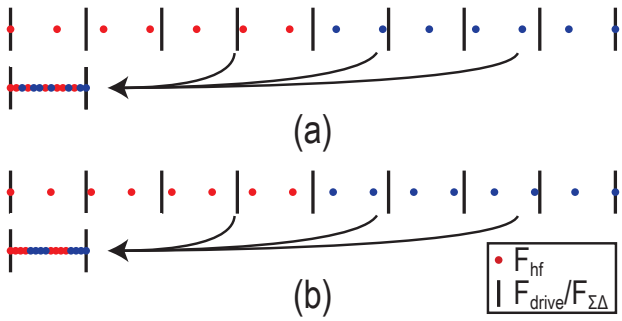

Figure 4.34: Interdigitation of F_{hf} phases.

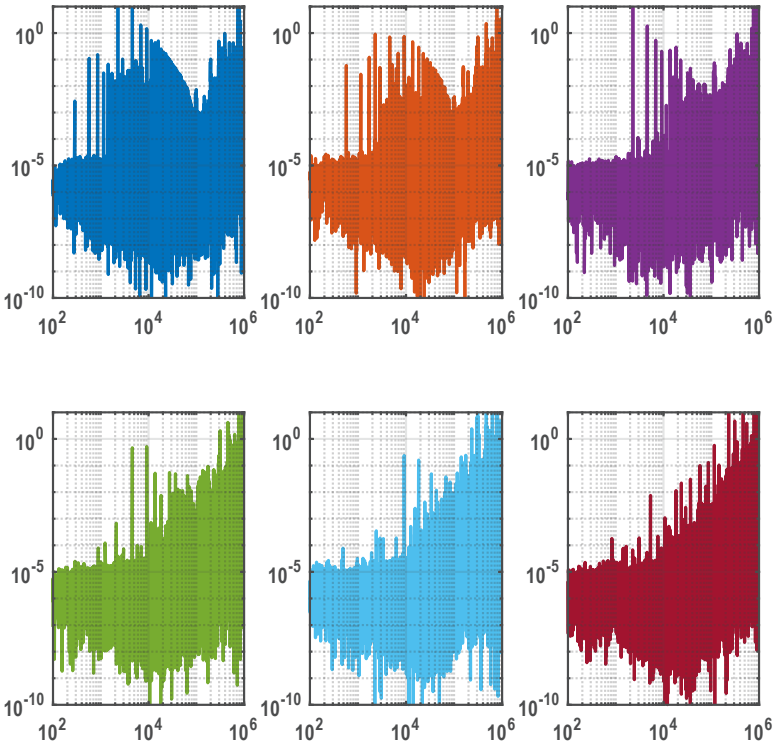

Figure 4.35: PSD of the bitstream for increasing interdigitation of F_{hf} clock phases.

Ideally, a beat period equal to the sampling period results in zero stochastic error from the time-domain induced dead zones. However, some of the dead zone-induced tones leak into the samples. This can be solved by moving these tones to a higher frequency by interdigitating the quantization noise pattern. The means by which these tones leak into the samples is illustrated in Figure 4.34, which is similar to Figures 4.23 and 4.24, but

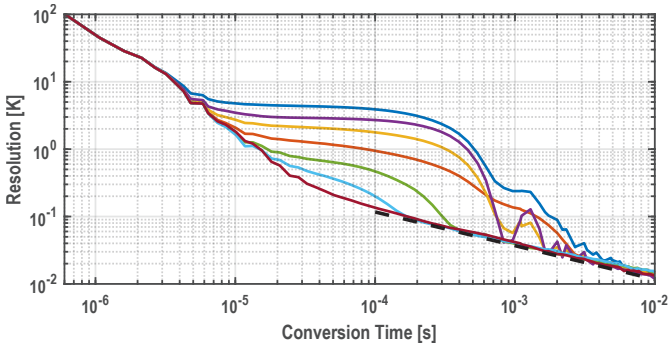

Figure 4.36: Resolution vs. conversion time for different levels of interdigitation of F_{hf} clock phases, but with the same beat period.

Table 4.4: Frequencies used in Figure 4.35 and 4.36.

F_{hf} [MHz]	M	F_{drive} [MHz]	N	Beat Period [ms]	M/N
570.004,577,6	622,597	1.875	2,048	1.09	304.002,441,4
555.004,577,6	606,213	1.875	2,048	1.09	296.002,441,4
562.504,577,6	614,405	1.875	2,048	1.09	300.002,441,4
566.254,577,6	618,501	1.875	2,048	1.09	302.002,441,4
568.129,577,6	620,549	1.875	2,048	1.09	303.002,441,4
569.067,077,6	621,573	1.875	2,048	1.09	303.502,441,4
569.535,827,6	622,085	1.875	2,048	1.09	303.752,441,4

now F_{hf} periods during the second half of the conversion are marked in blue. This allows us to visualize how the TQ-noise phase error behaves over the complete conversion. In (b), the second half of the conversion (blue dots) precisely detects the phase ranges where the blue dots are centralized, but creates large errors where they are not, i.e., during one half of the conversion TQ-errors are small while in the other half they will be large. As a result, the TQ-noise starts to correlate and becomes tonal. Even worse, in the case of (b), the tones will be located at low frequencies in the conversion spectrum. Alternatively, in (a), the TQ-noise phase errors are more randomly dispersed, which decorrelates the TQ-noise and spreads the related tones over a wider frequency band and to higher frequencies. As a result, they are hidden in the noise-shaped quantization noise and flat thermal noise (Figure 4.35) and are easier to filter out by the sampling window (Figure 4.36). The frequencies used in these plots are shown in Table 4.4. No measure exists that represents the level of dispersion of the TQ phase errors (i.e., the level of interdigitation between red and blue dots). However, the M/N ratio gives a good indication, because the further it is from being an integer or being a ratio with common divisor ratios, the

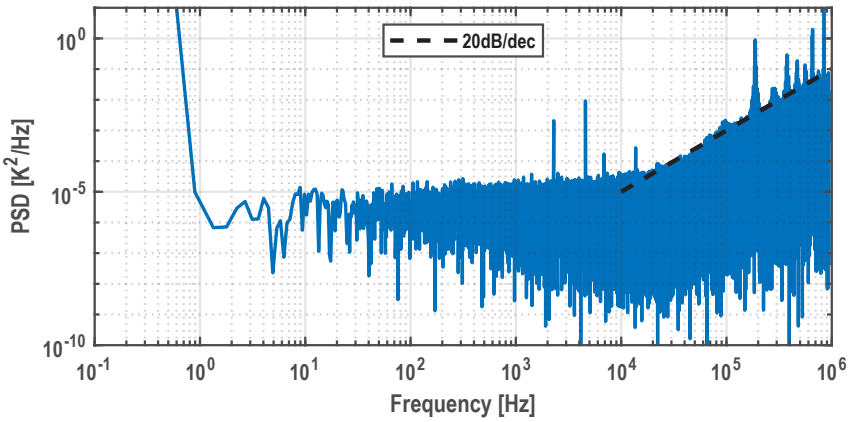

Figure 4.37: Bitstream PSD of 4,194,304 samples, Hanning-windowed.

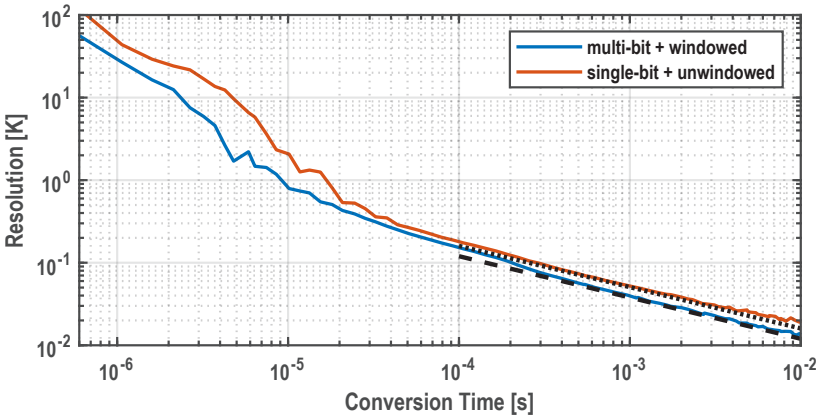

Figure 4.38: Resolution vs. conversion time of two operating modes.

better it tends to be. To improve the layout of the plots in Figure 4.35, it does not contain the measured data from row 3 in Table 4.2. Note that for this measurement, the beat period is fixed.

Figure 4.37 shows the PSD of the bitstream. It shows 1st-order noise shaping and no 1/f noise for the measured spectrum. It is quite tonal, which is expected for a 1st-order ΣΔ modulator. Figure 4.38 shows the resulting resolution vs. conversion time for two operating modes (multi-bit with windowing, and single-bit without windowing). The theoretical noise floor due to the TQ noise is shown for the windowed case (dashed line)

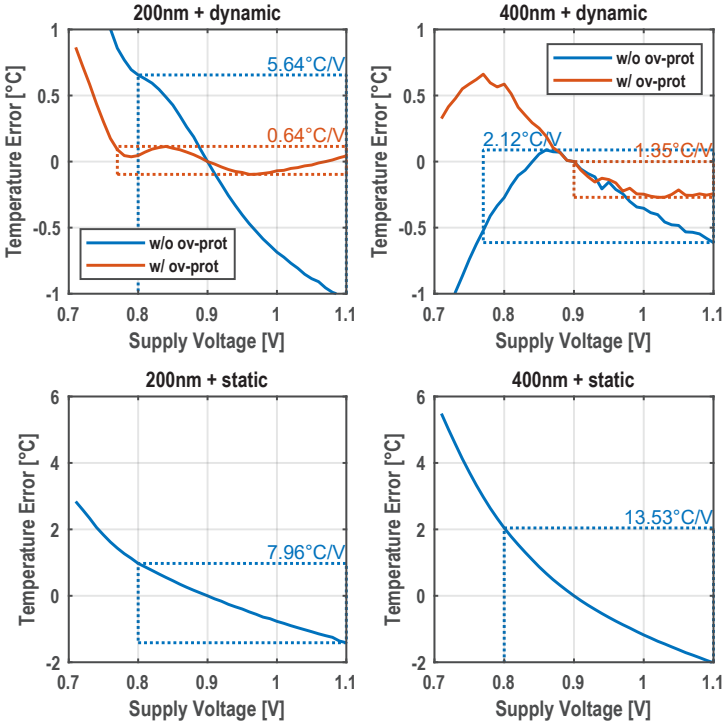

Figure 4.39: Measured power supply sensitivity.

and the unwindowed case (dotted line). Both modes perform according to expectations and achieve 40.6 mK and 53.2 mK resolution in 1 ms, respectively.

4.6.3 – DC Supply Sensitivity

Figure 4.39 shows the measured power supply sensitivity for the different versions. Since the dynamic comparator has overvoltage protection capability, these versions are measured with this feature both enabled and disabled. As expected, the overvoltage protection helps flatten the sensitivity. However, the effect is most pronounced in the 200 nm version. This is due to the larger PPF time constant, which makes other time-delay effects less significant. Moreover, its larger resistance means that the reduction in gate leakage due to the overvoltage protection is more significant compared to the resistor current. The technology's nominal supply voltage is 1.0 V, so in a voltage range of +10%/-23%, the sensor would exhibit an error less than +0.11°C/-0.10°C. This error is much smaller than the sensor's expected inaccuracy, which is a desirable property in thermal management applications.

4.7: Conclusion

This chapter describes the design of a compact resistor-based temperature sensor. Compared to a previous design (Chapter 3), this design has two main goals: 1) to reduce the complexity and reliance on analog circuitry, and 2) to remove the CCO from the signal path thus improving the accuracy. The first and most important goal is achieved: the sensor only requires two input frequencies. As a downside, the sensitive clocking scheme (section 4.5) requires a fractional PLL. However, sophisticated PLLs are readily available on SoCs [10,11,12] as they are required to implement DVFS. Furthermore, the system does not require an analog current reference, and the only analog node (the PPF's output) is immediately digitized by a fully differential dynamic comparator that is shown to be supply-independent. This means that the readout system is not only very small and fully synthesizable, but that it is also robust so that it can be placed in dense digital circuitry.

The major downside is the PPF, which acts as an all-pass filter for signals coming from the power supply. The ringing in the supply bond wires combined with the PPF's all-pass property causes oscillations in the phase readout, limiting the accuracy. Consequently, the PPF should not be considered in noisy enivronments (where adequate filtering/isolation is not possible) but should be replaced by a filter that incorporates a LPF, e.g., a WB or LPF.

4.8: References

[1] J. A. Angevare and K. A. A. Makinwa, "A 6800-μm² Resistor-Based Temperature Sensor With ±0.35 °C (3σ) Inaccuracy in 180-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 10, pp. 2649-2657, 2019.

[2] S. M. Kashmiri, S. Xia, and K. A. A. Makinwa, "A Temperature-to-Digital Converter Based on an Optimized Electrothermal Filter," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 7, pp. 2026-2035, 2009.

[3] S. Pan and K. A. A. Makinwa, "Optimum synchronous phase detection and its application in smart sensor interfaces," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, Baltimore, MD, 2017.

[4] U. Sönmez, F. Sebastiano, and K. A. A. Makinwa, "Analysis and Design of VCO-Based Phase-Domain ΣΔ Modulators," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 5, pp. 1075-1084, 2017.

[5] W. Choi, Y. Lee, S. Kim, J. Jang, J. Chun, K. A. A. Makinwa, and Y. Chae, "A Compact Resistor-Based CMOS Temperature Sensor With an Inaccuracy of 0.12 °C (3σ) and a Resolution FoM of 0.43 pJ · K² in 65-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 12, pp. 3356-3367, December 2018.

[6] Y. Lee, W. Choi, T. Kim, S. Song, K. A. A. Makinwa, and Y. Chae, "A 5800-μm² Resistor-Based Temperature Sensor With a One-Point Trimmed Inaccuracy of ±1.2 °C (3σ) From −50 °C to 105 °C in 65-nm CMOS," *IEEE Solid-State Circuits Letters*, vol. 2, no. 9, pp. 67-70, 2019.

[7] D. Schinkel, E. Mensink, E. Klumperink, E. van Tuijl, and B. Nauta, "A Double-Tail Latch-Type Voltage Sense Amplifier with 18ps Setup+Hold Time," in *IEEE International Solid-State Circuits Conference*, San Francisco, 2007.

[8] M. van Elzakker, E. van Tuijl, P. Geraedts, D. Schinkel, E. A. Klumperink, and B. Nauta, "A 10-bit Charge-Redistribution ADC Consuming 1.9 μW at 1MS/s," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 5, pp. 1007-1015, 2010.

[9] *I²C-Programmable Any-Rate, Any-Output Quad Clock Generator*, Si5338, Silicon Labs. [Online]. Available: `https://media.digikey.com/pdf/Data%20Sheets/Silicon%20Laboratories%20PDFs/Si5338.pdf`

[10] T. Singh, S. Rangarajan, D. John, R. Schreiber, S. Oliver, R. Seahra, and A. Schaefer, "Zen 2: The AMD 7nm Energy-Efficient High-Performance x86-64 Microprocessor Core," in *IEEE International Solid-State Circuits Conference*, San Fransisco, 2020.

[11] S. Naffziger, K. Lepak, M. Paraschou, and M. Subramony, "AMD Chiplet Architecture for High-Performance Server and Desktop Products," in *IEEE International Solid-State Circuits Conference*, San Fransisco, 2020.

[12] S. M. Tam, H. Muljono, M. Huang, S. Lyer, K. Royneogi, N. Satti, R. Qureshi, W. Chen, T. Wang, H. Hsieh, S. Vora, and E. Wang, "SkyLake-SP: A 14nm 28-Core xeon® processor," in *IEEE International Solid-State Circuits Conference*, San Fransisco, 2018.

Chapter 5 | Design III: Dual-GRO PDΣΔM

This chapter presents the design and measurement of a highly digital 2210 μm² resistor-based temperature sensor in 65 nm. Its main novelty is the use of a Dual-Gated Ring Oscillator (Dual-GRO), which will be shown to solve the effects of time-domain quantization noise (TQ noise). Despite its small size, it achieves a 1.3°C (3σ) inaccuracy from -55 to 125°C after a correlated 1-pt trim. Thanks to its novel and highly digital readout, it achieves a good power, resolution, and area trade-off.

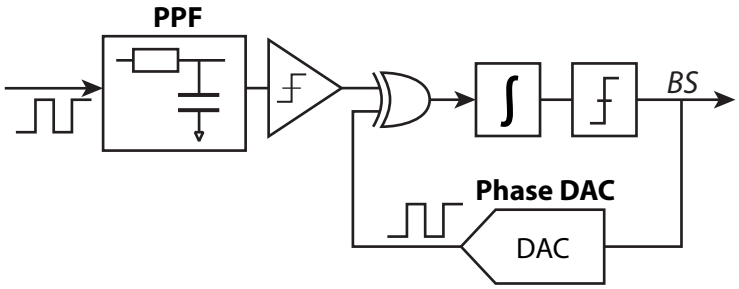

Figure 5.1: Simplified system diagram.

The design presented in the previous chapter is limited by time-domain quantization noise (TQ noise) in two ways. First, TQ noise limits the sensor's resolution. To minize this effect, the counter's clock frequency can be increased, but this comes at the expense of increased power consumption. The final result is that, TQ noise limits both resolution and power consumption. Second, it influences the sensor's resolution through the interplay between the counter clock frequency and the driving frequency. The resulting limit cycles may cause large dead zones in the sensor's readout (up to 10°C). Careful selection of the counter frequency with a high-performance fractional PLL is required to mitigate this problem.

The main goal of this chapter is to design an improved readout which removes the counter's TQ noise, thereby removing the limit-cycling problem of the previous design and its need for a fractional PLL. Furthermore, by removing the TQ noise, the counter frequency vs. resolution trade-off is eliminated, improving the system's FoM by a factor of 50x. Even though some increase in complexity is required, the secondary goal of this design is to minimize this increase while maintaining the previous design's simplicity as much as possible.

Figure 5.1 shows a simplified system diagram of the new design. The PPF, static comparator, and linearized XOR-mixing scheme are reused from the previous design. The main difference lies in the design of the integrator, which uses a novel dual gated ring oscillator (GRO) that does not suffer from time-quantization noise (see Section 5.2). To explain how the dual-GRO removes this noise, first, a new insight into time-quantization noise must be given (see Section 5.1).

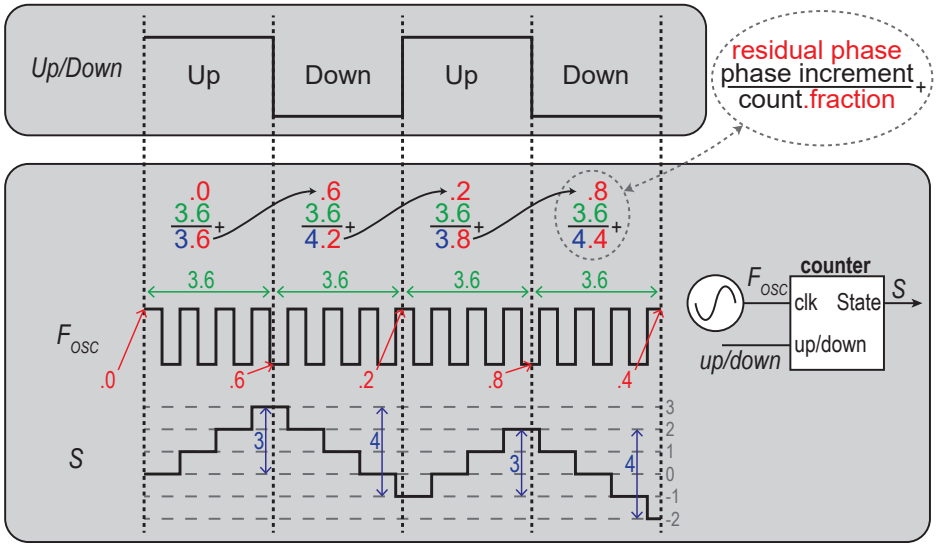

Figure 5.2: Simplified example illustrating the origin of TQ noise.

5.1: Time Quantization Noise

The main cause of the TQ noise is that part of the ΣΔM's integration-state is stored in the counter, while the fractional part is stored in the oscillator's phase. While the polarity of the counter part of the integration state is toggled (chopped) when the up/down signal is toggled, the fractional state stored in the oscillator's phase is not. Thus, the oscillator's fractional state keeps on changing regardless of the up/down cycles, starting and stopping at seemingly random phases at the beginning and end of each up/down cycle.

This is shown in the simplified illustration in Figure 5.2, where the counter integrates an up/down signal with a 50% duty cycle. Ideally, this should result in a net integrated value of zero. In the example, each up and down period causes an increase in oscillator state of 3.6 cycles. Since the counter cannot count fractional numbers, it only stores the value 3, missing the 0.6 cycles that are stored in the oscillator. What exacerbates this problem is that the oscillator starts at 0.6 cycles in the next down-period, resulting in a subtraction of 4 counts from the counter state. This rounding effect happens every up/down-period, resulting in a count error of -2 at the end of the two up/down-periods.

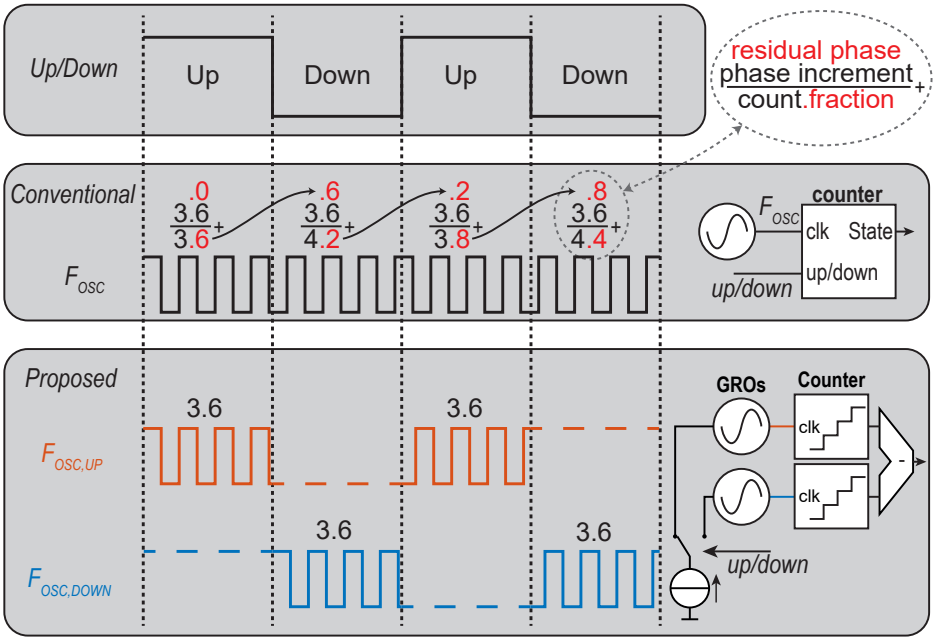

Figure 5.3: Simplified example demonstrating how TQ noise is resolved.

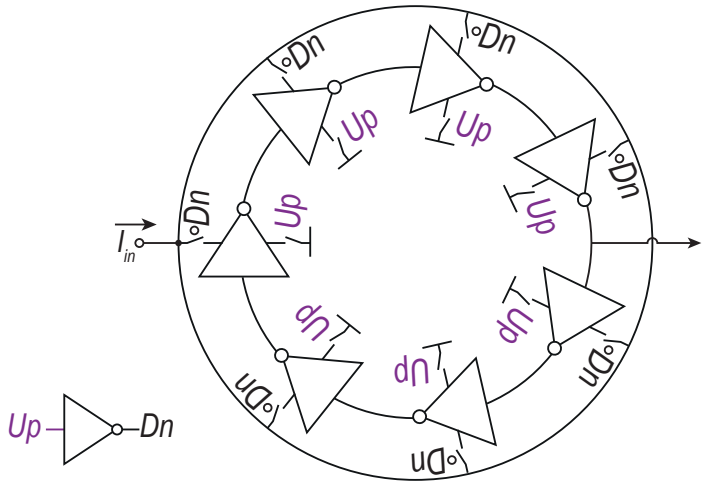

Figure 5.4: Simplified diagram of a GRO.

The counter's TQ noise can be suppressed by locking the oscillator state to the counter state, i.e., by keeping the fractional component. This requires a separate up and down oscillator and counter pair, as shown in Figure 5.3. In the example, the oscillator state (and counter state) can be frozen by using

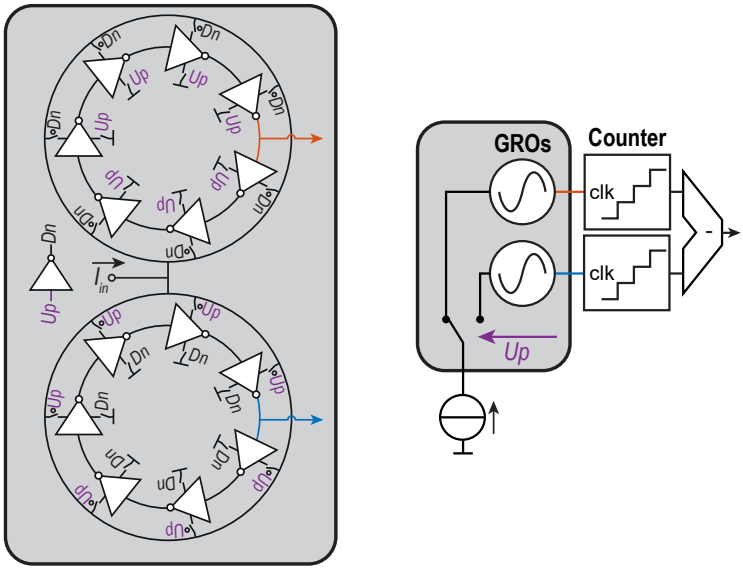

Figure 5.5: Simplified diagram of the dual-GRO.

a GRO [1]. After the locking period, the oscillator and counter continue where they remained from the previous cycle. As a consequence, there are now two separate states (up and down) which are combined through digital subtraction. Note that this operation occurs in the digital domain, which provides both simplicity and scalability.

The GRO is a ring oscillator that can be gated. To achieve this, each stage of the ring oscillator can be enabled/disabled (see Figure 5.4). When the GRO is enabled, it behaves exactly like a normal RO. However, when disabled, no charge can enter or leave the inverter nodes, and so its state is frozen.

5.2: Dual GRO

To count both up and down, separate GROs are required (see Figure 5.5). More importantly, the operation of the GROs should be mutually exclusive, with one of the two GROs always enabled. To minimize the complexity of controlling switches, the two GROs share the same enable/disable signals, but with complementary inputs. Since a disabled GRO does not draw current from its supply, the supply current is directly connected to both GROs and automatically finds its way into the enabled GRO. In the following subsections, different types of GRO imperfections will be presented, and their effect on the sensor's performance will be analyzed.

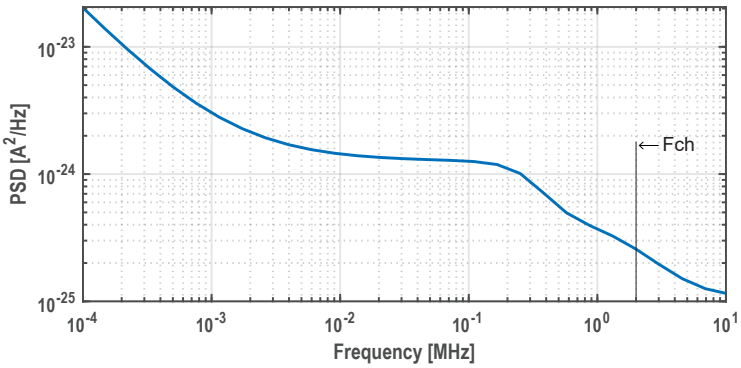

Figure 5.6: Bias current noise that is on top of a ~1 µA bias current. Most of its energy is below the mixing frequency (F_{ch} which is equal to F_{drive}, ~2 MHz), and is filtered out.

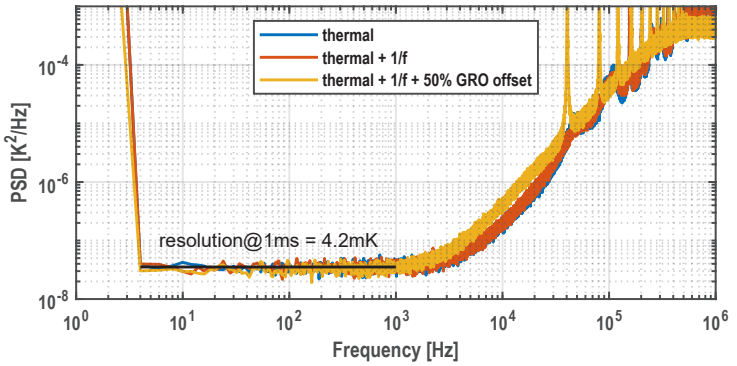

Figure 5.7: PSD of the bitstream where only the bias current contains noise.

5.2.1 – Current Bias

To operate at a desired frequency, the dual GRO requires a bias current which is readily available in most SoCs for the purpose of driving other analog circuits. A central bias block generates and distributes the bias currents to all the temperature sensors.

The integrator in a PDΣΔM is insensitive to 1/f-noise since it is a chopped system [2]. In the same way, this phase readout is insensitive to GRO frequency variations, and therefore also to bias current variations. The intuitive explanation why frequency variations, drift, and 1/f noise do not

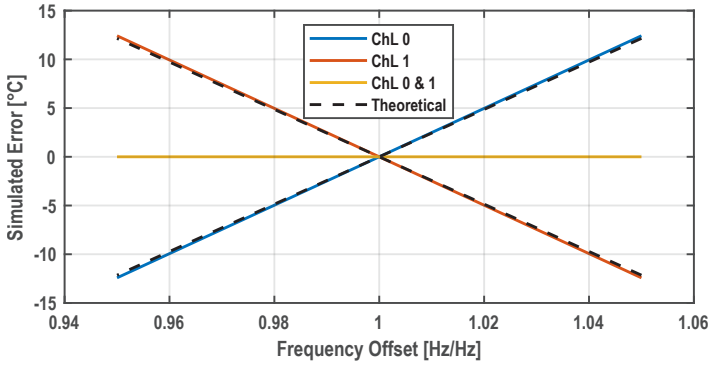

Figure 5.8: Measured error due to frequency offset between the two GROs.

influence the sensor is that these changes in frequency affect both GROs. As a result, the only difference is the signal swing in the integrator, while the output is still perfectly proportional to the duty cycle of the up/down signal.

Figure 5.6 shows the current noise spectral density of the bias current that is on top of a ~1 µA current. Noise that is below the demodulation frequency is chopped by the mixer/demodulator of the PDΣΔM and therefore does not influence the resolution. This is shown in Figure 5.7, which shows the bitstream noise floor with and without bias 1/f noise. It shows that the noise floor remains the same, irrespective of the presence of 1/f-noise. Moreover, even with a frequency error between the GROs of 50%, the in-band noise due to 1/f-noise remains unchanged, although it does show the up-modulated 1/f-noise more clearly.

5.2.2 – Offset

While common-mode frequency variations in both GROs do not affect the readout, differential variations do. With unmatched GROs, the integrated value is not proportional to the duty cycle of the up/down signal, but has an offset that is proportional to the frequency difference. This then results in a phase readout with an offset that is proportional to the frequency difference, as shown in Figure 5.8. This matches well with the theoretical error, which is:

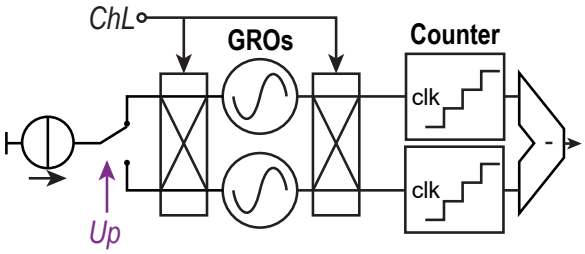

Figure 5.9: System-level chopping (ChL) of the dual-GRO.

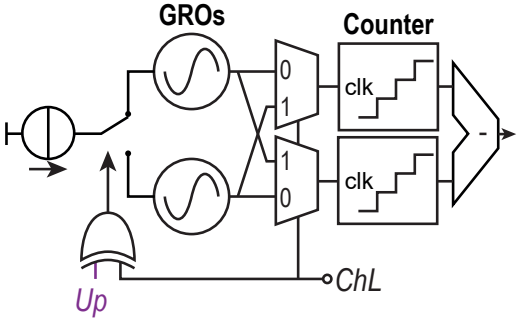

Figure 5.10: Synthesized digital implementation of ChL.

$$T_e = \alpha \cdot 90° \cdot \frac{\Delta F_{CCO}}{2 \cdot F_{CCO}} \quad (5.1),$$

where $\alpha \approx 5.4$ °C/° is the PPF's sensitivity.

As in any system with mismatch, the error can be mitigated by swapping (chopping) the two GROs so that differential low frequency errors can be filtered out (see Figure 5.9). As shown in Figure 5.8, the low frequency variations between the two GROs can be rejected after system-level chopping (ChL). The only residual effect from the offset is a reduced maximum phase range. The simulated 6σ GRO frequency error is 2%, and so the loss in phase range is less than 1°. Finally, to minimize complexity and maximize scalability, the ChL is implemented in fully synthesized digital logic (see Figure 5.10).

A downside of using ChL is that it severs the link between the counter and the oscillator. As a result, it causes TQ noise, i.e., the noise that the dual GRO is meant to eliminate. However, in this case, the TQ noise is injected only once every full converion instead of once every ΔΣM cycle. As a result, the inband TQ noise becomes:

$$\sigma_{T,IB} = \alpha \cdot \sigma_{\Delta C} \cdot \frac{90^{\circ}}{F_{GRO}} \cdot \sqrt{\frac{F_{ChL}}{T_{conv}}} \quad (5.2),$$

where $\sigma_{T,IB}$ is the in-band resolution in Kelvin, α is the temperature-to-phase sensitivity, $\sigma_{\Delta C}$ is the standard deviation of the integrated error at every chop transition (and is $\sqrt{2}$ in this case; see below), F_{GRO} is the GRO's frequency, F_{ChL} is the system level chopping frequency, and T_{conv} is the total conversion time.

Again, the integrated error $\sigma_{\Delta C}$ is different from previous cases and is found as follows: at the beginning of a conversion, both up and down GROs start at a random state (uniform distrubtion between 0 and 1 cycles). When the GROs are swapped, the GRO phase that was in the up-state is replaced with the other GRO's phase. This results in an error in the up-state that has a uniform distribution between -1 and 1. The down-state gains the same error but of inverse polarity, i.e., they are correlated. Therefore:

$$\sigma_{\Delta C} = \sqrt{\frac{1}{12} \cdot [1^2 + 1^2 + (2+2)^2]} = \sqrt{\frac{3}{2}} \quad (5.3).$$

On top of that, swapping the GROs can cause one of the counters to experience a rising edge which causes it to increment its count. For this to happen, one GRO output should be low, while the other should be high at the swapping moment. The chance of this happening depends on the duty-cycle D:

$$P(+1) = D \cdot (1 - D)$$
$$P(0) = D^2 + (1 - D)^2$$
$$P(-1) = (1 - D) \cdot D \quad (5.4),$$

As a result, its variance is:

$$\sigma^2 = P(+1) \cdot 1^2 + P(0) \cdot 0^2 + P(-1) \cdot -1^2 = 2 \cdot (D - D^2) \quad (5.5),$$

which for a 50% duty-cycle is ½. The error due to a counter incrementing its count is expected to correlate, to a certain degree, with the error due to the GRO state swapping. However, when the integrator swing is large, this should converge to random behavior, and so Eq. (5.3) is adapted to include this clocking error:

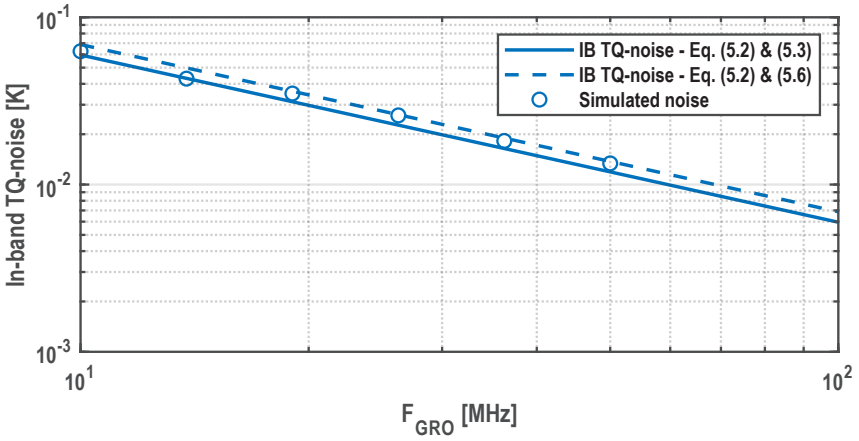

Figure 5.11: In-band TQ noise over F_{GRO}, using eq. (5.4) & (5.5), where T_{conv} = 1 ms and F_{ChL} = 1 kHz, compared to simulated values.

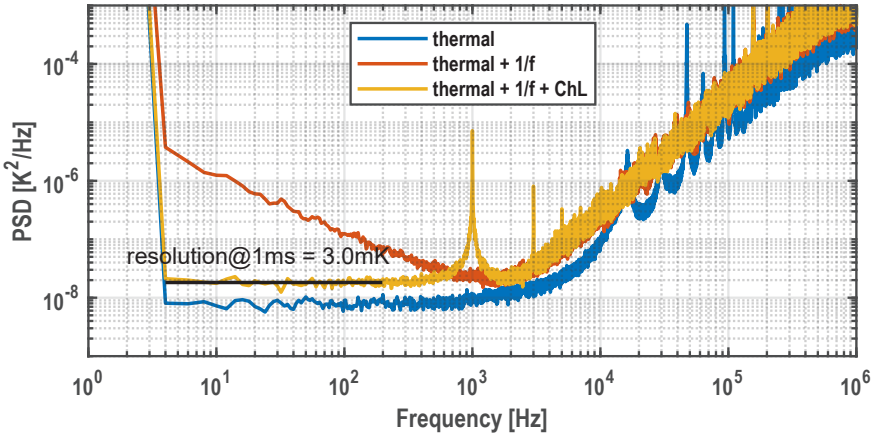

Figure 5.12: PSD of the bitstream where only the GRO contains noise.

$$\sigma_{\Delta C} = \sqrt{\frac{1}{12} \cdot [1^2 + 1^2 + (2+2)^2] + \frac{1}{2}} = \sqrt{2} \qquad (5.6).$$

The resulting in-band noise is plotted in Figure 5.11, for T_{conv} is 1 ms and F_{ChL} is 1 kHz (dashed line uses ${\sigma_{\Delta C}}^2 = 2$; solid line uses ${\sigma_{\Delta C}}^2 = 3/2$).

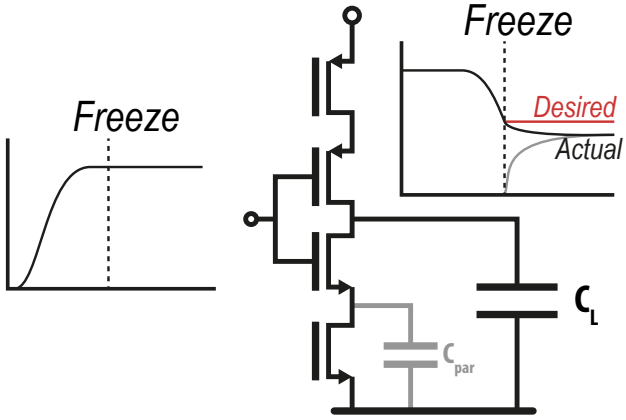

Figure 5.13: GRO state corruption after freezing the GRO.

5.2.3 – Phase Noise

ROs are known to contain substantial phase noise, but offer a small footprint in return. The same applies to GROs, and this noise will still limit the sensor resolution. This can be minimized by increasing power consumption, but that comes at the cost of area. The voltage drop across the RO must be small enough so that the current source supplying its power remains saturated over PVT. Therefore, resizing is necessary when the power is increased.

Figure 5.12 shows how GRO noise affects the bitstream PSD for three different cases: 1) where there is only thermal noise from the GRO and the system level chopping is disabled; 2) where only the GRO contains thermal and 1/f noise and system level chopping is disabled; and 3) where only the GRO contains thermal and 1/f noise, and the system level chopping is enabled. This shows that 1/f noise in the GRO is effectively removed by ChL.

5.2.4 – Parasitic Charge

A less apparent source of corruption in the GRO is the mixing of parasitic charge when the GRO is frozen. The GRO stage that transitions when the GRO is frozen still has some charge remaining on its output, while the parasitic capacitance that is shown in Figure 5.13 is empty just before the GRO freezes. This parasitic charge and the stage's output charge are then mixed together, corrupting the GRO state by moving the phase forward, depending on its phase during the freezing. This dependence is shown in

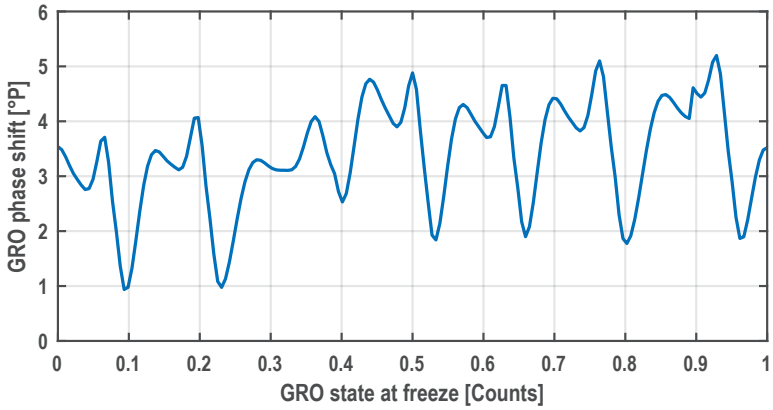

Figure 5.14: State-dependent GRO phase-shift due to parasitic capacitance.

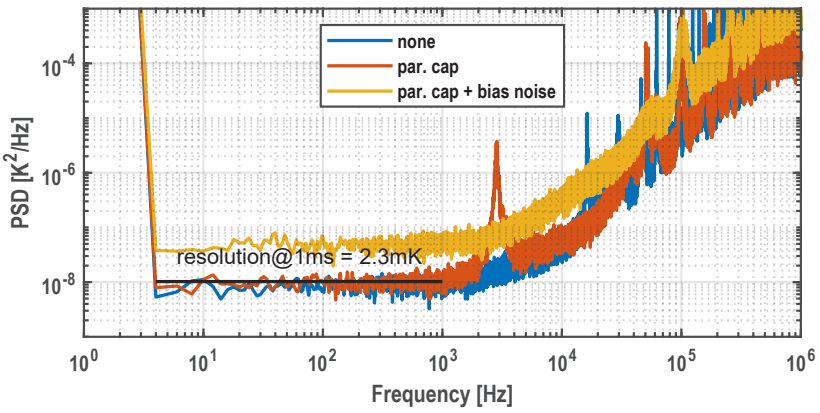

Figure 5.15: PSD of the bitstream with GRO state corruption due to internal parasitic charge.

Figure 5.14, which shows the phase increase of the GRO depending on its state at the freezing moment. There are 14 transitions in an RO of 7 stages. Since the 7 falling edges discharge over a low ohmic VSS while the 7 rising edges charge from a high-impedance current source, their effect differs in magnitude. The mean effect, to the first order, is removed by both GROs having the same mean effect. However, if the moment of freezing is random, then the phase corruption is random. The 1σ deviation of this phase corruption $\sigma_{PC} \approx 1°$, and as a result its in-band noise is:

$$\sigma_{T,IB} = \alpha \cdot \frac{2 \cdot \sigma_{PC}}{360°} \cdot \frac{90^{\circ}}{F_{CCO}} \cdot \sqrt{\frac{F_{drive}}{T_{conv}}} \quad (5.6),$$

which approximates to 2.4 mK.

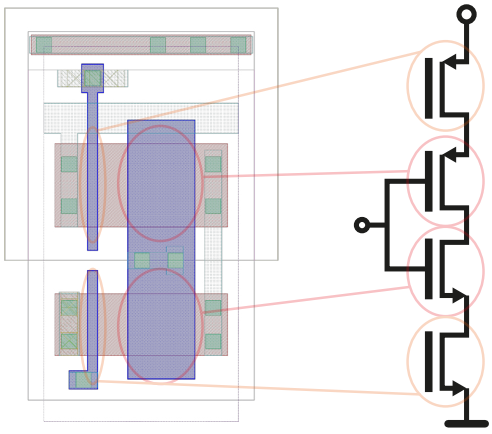

Figure 5.16: Layout of the GRO inverter stage which minimizes the parasitic capacitance compared to the loading capacitance (blue = gate, red = diffusion).

Figure 5.15 shows the PSD of the bitstream under three conditions: 1) only PPF and comparator noise are present, 2) the GRO parasitic capacitance effect is also present, and 3) on top of that bias noise is also present. The PPF and comparator noise are added to this simulation because the state corruption due to parasitic charge is fully deterministic. This can lead to limit-cycling behavior [3], which is visible when the effect of the parasitic cap is enabled. However, after adding noise to the bias current, the freeze phase corruption is sufficiently randomized.

The effect of the parasitic charge of the GRO can be minimized either by maximizing the stage capacitance relative to the parasitic capacitance, or by maximizing the GRO frequency, which maximizes the integrator swing so that the phase corruption has a smaller relative effect. Increasing the GRO frequency increases power consumption and increases the counter/subtraction bit-size requirement to prevent overflow. Increasing the relative capacitance can be done by maximizing the area of the RO stages. An optimum layout ensures that the diffusion of the enable switches is connected through the shortest possible diffusion length. This removes parasitic capacitances due to the interconnect and only requires a single diffusion, which acts as drain and source for the two transistors (see Figure 5.16). The big dark-blue rectangle on the left represents the gates of the main inverters, while the smaller dark-blue lines that are on the right are the enabling switches.

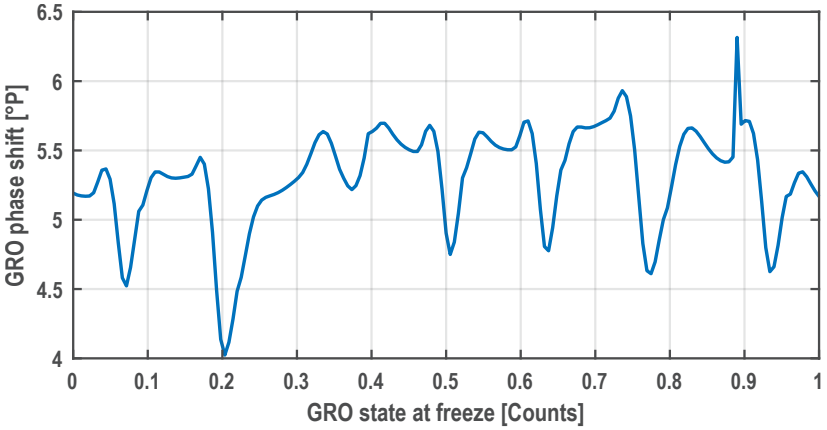

Figure 5.17: GRO state corruption due to 100 ns of switch leakage.

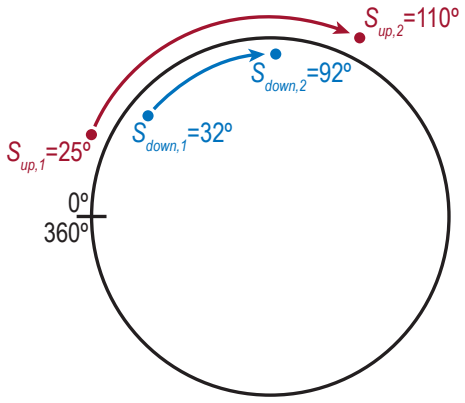

Figure 5.18: The up-GRO (S_{up}) overtaking the down-GRO (S_{down}), without a clocked difference in the counter, resuling in an effective delay.

5.2.5 – Switch-Leakage

When open, the freeze switches of the GRO leak, especially at higher temperatures. This results in a non-zero freeze current that is integrated. To first order, both GROs suffer from the same leakage current, which, just like bias current variations, causes an increase in the integrator state without affecting the proportionality to the up/down duty-cycle. What is more, even differences in leakage current between the GROs are filtered out by the *ChL*.

Leakage can have a varying effect depending on the state of the GRO, and as a consequence, can cause stochastic errors. Figure 5.17 shows the GRO phase corruption under worst-case conditions (PVT) after 100 ns

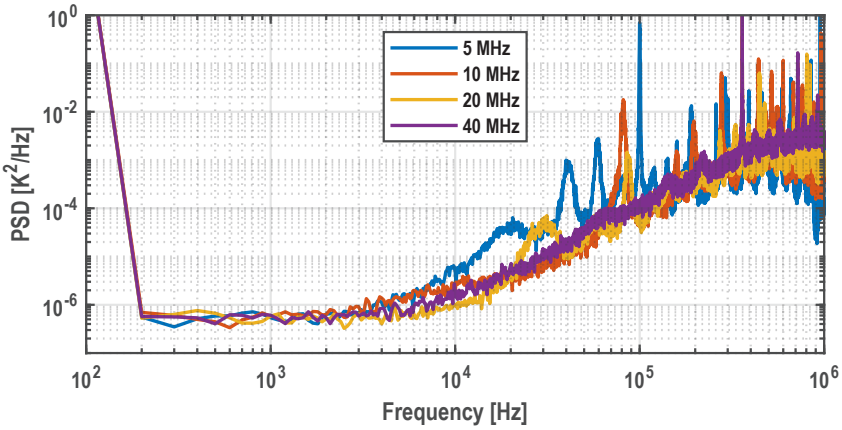

Figure 5.19: PSD of simulated bitstream for different GRO frequencies showing unstable and tonal behaviour at lower GRO frequencies.

of freezing time. It shows similar behavior to the GRO corruption due to its parasitic charge. However, it shows a slightly larger mean effect, which is filtered out by *ChL*, while also showing a much smaller varying effect, which causes even less noise than the GRO's parasitic charge.

5.2.6 – Excess Loop Delay

Now that the TQ noise problem of the counter integrator has been solved, the main question is: how far can we minimize power consumption by reducing the GRO frequency? At first glance, it is logical to assume that, since the counter still has a quantized view of the complete integrator state, the limit is determined when this source of quantization dominates the ΣΔM's normal quantization noise. However, before this happens the ΣΔM will suffer from an effect similar to excess loop delay (ELD). Figure 5.18 shows an intuitive explanation. Suppose that the GRO frequency is much lower than the ΣΔM sampling/driving frequency. Between two sampling moments, one GRO overtakes the other GRO, which means that the bistream should toggle. However, the counters do not register a difference and so the ΣΔM does not immediately detect this change in integration state. Note that this also happens when both GROs pass the clocking phase (0°) the same number of times. This can lead to instability and tonal behavior in the ΣΔM, as shown in Figure 5.19. As the GRO frequency relative to the driving frequency is increased, the integrator swing increases, therefore the chance of the ΣΔM missing an overtake between the two GROs is reduced.

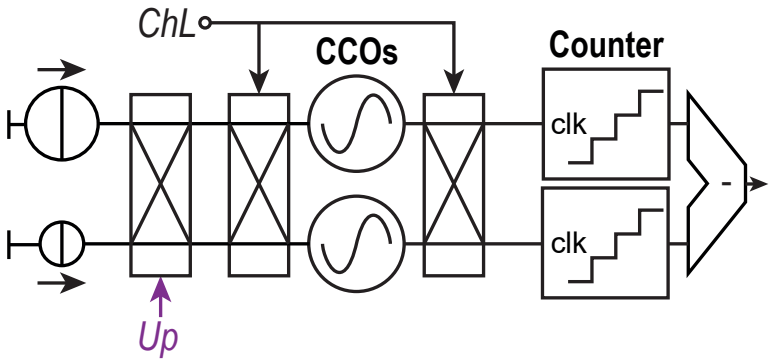

Figure 5.20: Dual-CCO implementation.

Strategies that compensate for the apparent loopfilter delay can be applied to resolve this phenomenon [4]. However, implementing these strategies is no trivial matter and steeply increases the complexity of the design. Moreover, the tonality is already inmeasurable at GRO frequencies above 50 MHz while the resulting digital power consumption is still very low (less than 2 µW).

5.2.7 – Dual-CCO

Solving the TQ noise problem by locking the oscillator state to the counter state does not neccesarily require a GRO. Alternatively, the RO that is frozen can be operated at a reduced frequency (see Figure 5.20). The major difference with this dual-CCO arrangement is that the integrator swing is reduced by this lower frequency since the integrator swing is proportional to the frequency difference between the two GROs. As a result, effects of the dual-GRO which are sensitive to integrator swing will be more pronounced unless the RO frequency is increased. This is similar to the leakage problem, where the leakage current of the GRO implementation acts as the reduced frequency. However, in that case the leakage current is so small that it barely has any effect on the integrator swing.

The dual-CCO implementation has two main downsides. The first is that the power consumption is increased. This is because both oscillators and counters are always counting, while the frequency difference should remain the same. The second, and larger concern, is the possibility of injection locking between the two ROs. Since the ROs require good matching to minimize the effect of some of the error sources, they are placed in close proximity. However, this facilitates injection locking. Furthermore,

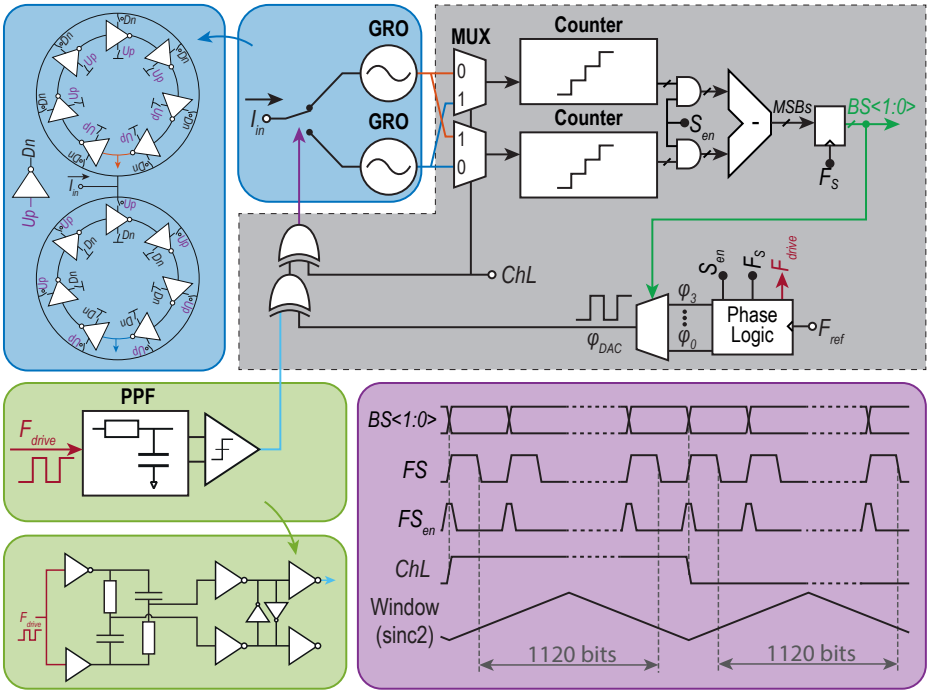

Figure 5.21: System diagram.

injection locking and its effects are not easily similated, since the substrate is not modeled in the simulator. As a consequence, the GRO implementation is preferred.

5.3: System

Figure 5.21 shows the complete sensor system. The PPF and static comparator are reused from the previous design (Ch. 4) to generate a digital signal whose phase shift is temperature-dependent. The PDΣΔM's mixer is then implemented by an XOR gate. Low-frequency chopping (*ChL*), completely implemented in synthesized digital logic, chops low-frequency errors from the dual-GRO. This occurs once every conversion. To minimize the errors due to the chopping transient, a sinc^2 window with its minima around the *ChL* transitions is used. Separate GRO and counter pairs count the up- and down-periods. The complete integration state is then obtained by applying a subtraction in the digital domain. Finally, a phase-logic block generates all the required phase and frequency signals from a reference frequency.

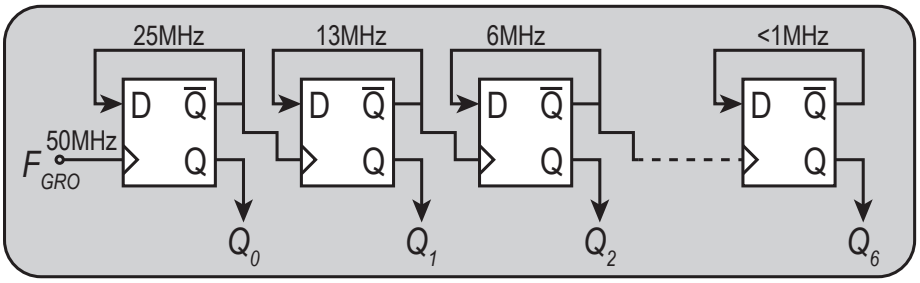

Figure 5.22: Ripple counter.

The main idea and innovation of this design is extensively described in previous two sections. However, some of the sub-systems described here require fine-tuning to improve system peformance. These are briefly described in this section.

5.3.1 – Comparator

The comparator is a crucial element to achieve good accuracy. Its effect on resolution is less important since ~mK of resolution is easily achieved [5,6] while there is no benefit to measuring below tens of mK. Thus, the comparator is mainly optimized for accuracy.

The presence of the dual-GRO means that the readout circuit is a continuous-time system. As a result, a dynamic comparator cannot be used since it would periodically sample the input signal. Consequently, a continuous-time comparator is required, which must then be carefully designed to minimize its offset, delay, and hysteresis. The first step is to start with minimum length input inverters and then to increase their width, while keeping the switching threshold at $0.5V_{DD}$ under nominal conditions, until their input impedance (due to gate leakage and input capacitance) starts to affect the accuracy. The next step is to use minimum size inverters for the cross-coupled inverters and increase their length until the comparator is just short of hysteresis under nominal conditions (see Section 4.3), while keeping their switching threshold at $0.5V_{DD}$.

5.3.2 – Counter

Prior-art highly digital PDΣΔMs require an up/down counter, while the improved PDΣΔM of this chapter only requires a counter that increments its count. This has two main advantages. First, while an up/down counter requires full adders to calculate its new state, a counter that only increments its count can be implemented as a ripple counter, which does not require any logic between the DFFs of the counter (see Figure 5.22). This reduces both area and power consumption. Second, every consecutive bit in a ripple counter runs at half the frequency of its previous bit, resulting in an automatic energy reduction in the counter. In previous up/down counters, the combination of a 2-bit gray-code counter and a traditional binary counter are used to achieve similar frequency and power scaling. However, this only achieves a one-time frequency reduction, while the ripple counter achieves the same for all of its bits. Furthermore, the counter segregation also comes at the price of increased complexity and digital gates. In contrast, this system has atomatic power scaling between each bit. One downside is that this new system requires two counters instead of one, although these counters require fewer bits since the GRO frequency is lower.

The counter size required to prevent overflow in the integrator is 7 bits. This is calculated using Eq. (2.26), where F_{CCO} = 100 MHz, F_{drive} = 2 MHz, and $\varphi_{in} - \varphi_{ref}$ = 45°, which is the maximum expected phase difference at the extreme corners. The resulting integrated count value of 25 must at least be smaller than ¼ of the integrator span, and so the total span required is 100 ≈ 7 bits. The actual worst-case phase difference that can occur because of bit error rate (BER) is 67.5°, but even in that case the ΣΔM can still handle GRO frequencies up to 85 MHz.

Another benefit of the dual GRO's counters is its relaxed meta-stability requirements. In this design there are three clock domains: one for each GRO, and an Fref. Clock signals that are generated in one clock domain, but sampled in another, can give rise to meta-stability issues. In this system, there are signals originating from Fref that cross to the GRO's clock domain. However, these signals are never clocked by the GROs and do not give rise to meta-stability, i.e., the GROs are analog in time. The only other clock-domain crossing signal is the subtraction result which is generated in the GRO clock domain and sampled by the Fref clock domain. Consequently, this can give rise to meta-stability. However, as long as the chance of meta-stability is small and the integrator does not overflow, this

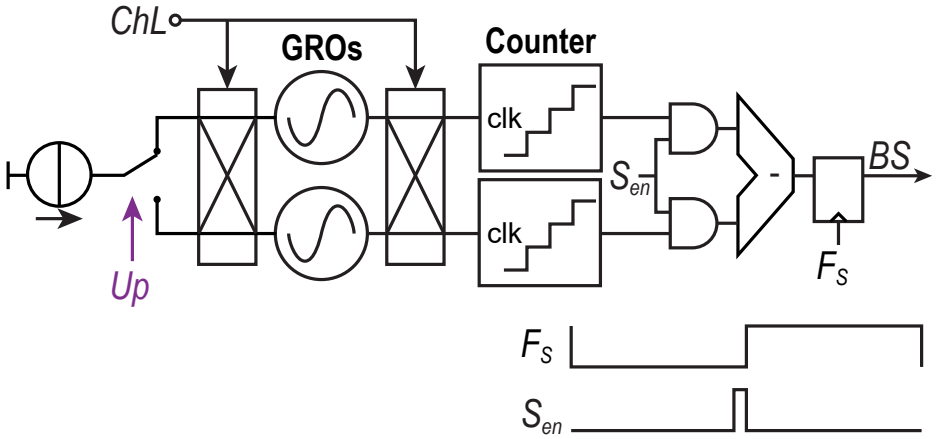

Figure 5.23: Enabled subtraction just before bitstream sampling.

effect only causes the bitstream to dither. As a result, the dual-GRO design is less sensitive to meta-stability, because even if it occurs, the integrator state is still maintained. The simulated meta-stibility chance is small enough that it is only expected to happen once every five conversions ($5 \cdot 1$ ms).

5.3.3 – Subtraction

Since one of the two counters is always counting at ~50 MHz, the subtraction bits also toggle at this frequency. This in turn consumes power continuously whereas the subtraction result is only required just before the bitstream is sampled. As a result, the subtraction node consumes ~9 µW, which is ~25% of the total power consumption. Therefore, the subtraction node is gated by AND-gates and only enabled just before the bitstream is sampled (see Figure 5.23). As a result, the power consumption is reduced to < 1 µW.

5.3.4 – Phase DAC

A phase DAC generates phases from a higher frequency clock so that all the phases can be generated using synchronous digital logic to achieve accurate and precise timing. Figure 5.24 shows the desired signals that the phase DAC generates. To generate phase increments of 22.5° at 2 MHz, a 32 MHz F_{ref} is used so that one of its periods corresponds to a 22.5° phase shift. This involves dividing Fref by 16 using 4 stages of a ripple counter structure (see Figure 5.25). The specific phase shifts are then generated by reclocking this 2 MHz signal by any of the signals in the frequency division chain which then cause phase shifts proportional to their period (see Figure

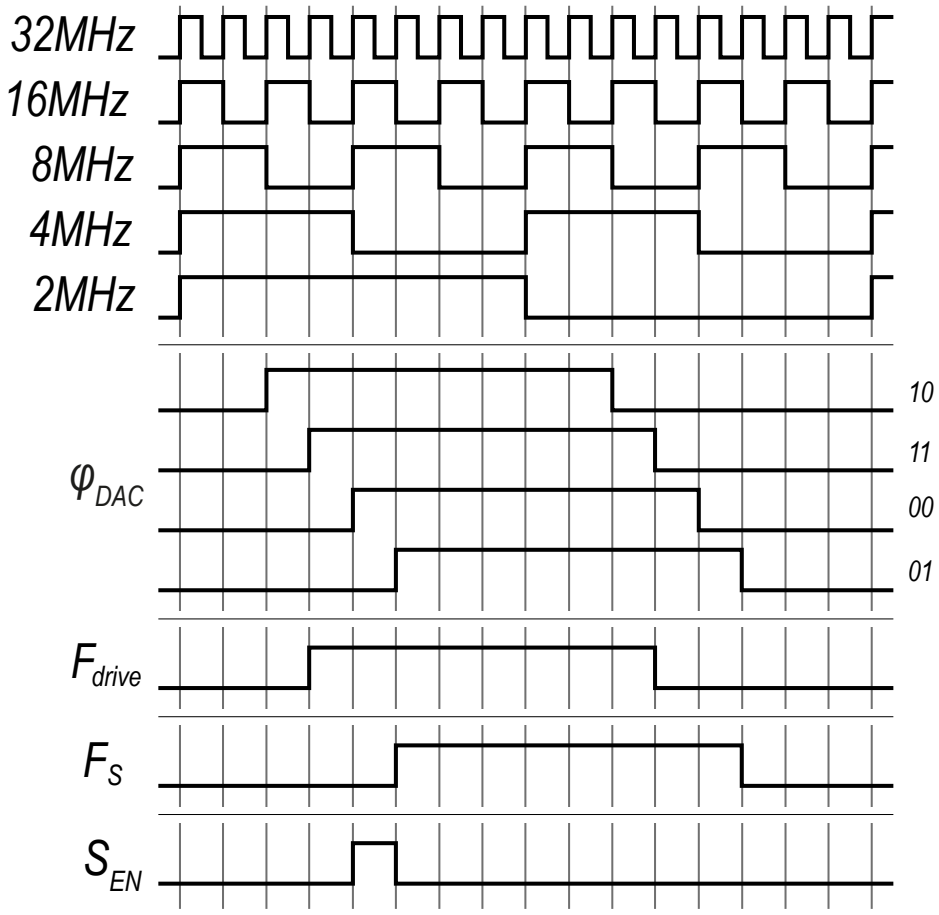

Figure 5.24: Desired signals that the phase DAC should generate.

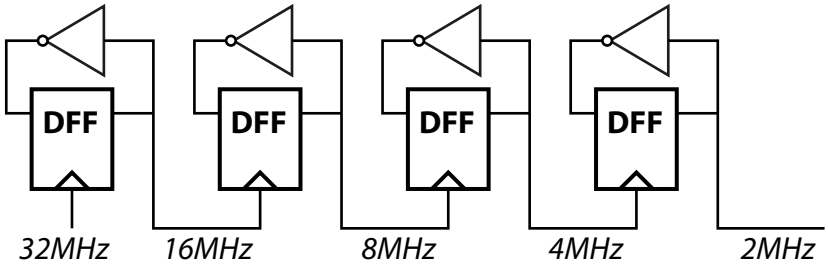

Figure 5.25: Phase DAC frequency division.

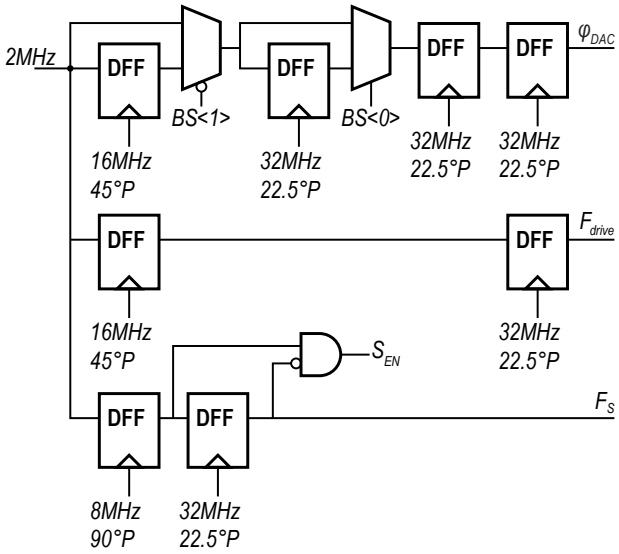

Figure 5.26: Classical implementation of a minimalized phase DAC.

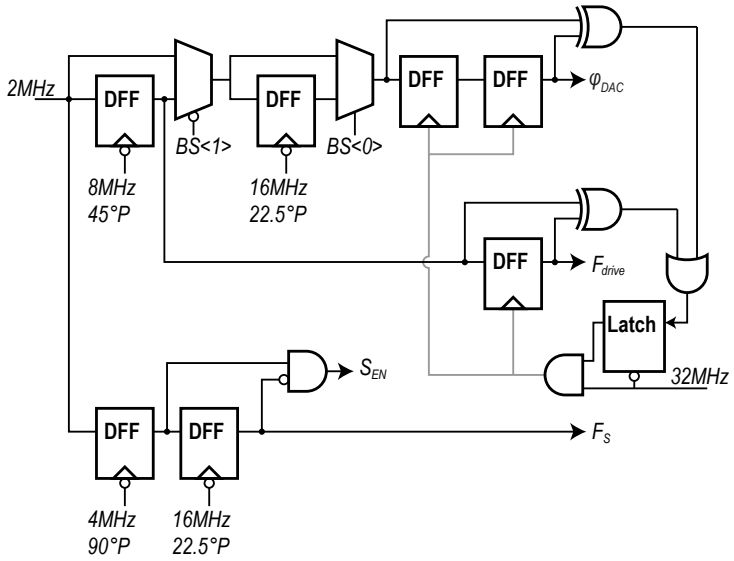

Figure 5.27: Optimized implementation of a minimalized phase DAC.

5.26). The phase delays that are controlled by the bitstream can be enabled/disabled by a multiplexer. The subtraction enable signal (S_{EN}) is enabled in the timeslot leading to the sampling of the bitstream. Finally, the input (F_{drive}) and reference phase (φ_{DAC}), whose relative phase delay is of vital importance, are reclocked using the 32 MHz F_{ref} signal, which does not contain clock-to-q timing skews.

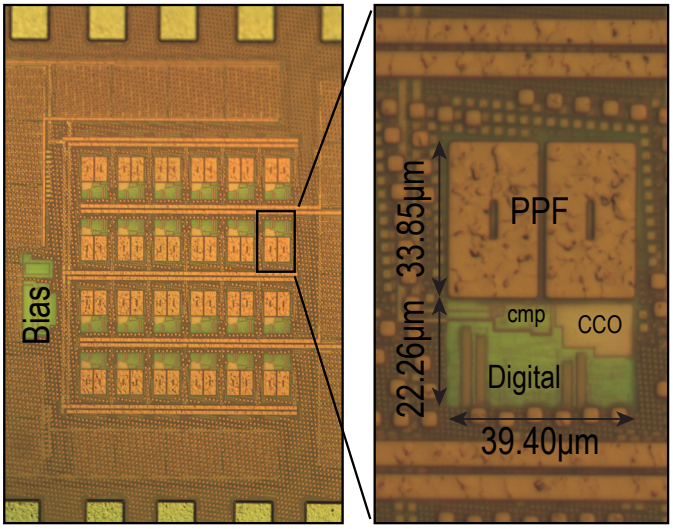

Figure 5.28: Chip photo.

The sensor design of this chapter aims to drastically decrease power consumption to the point that a traditional phase-DAC design will impact the power consumption. Therefore, two power saving techniques are applied to this phase DAC (see Figure 5.27). The first is based on the observation that the phase shifts can be generated by clocking on the falling edge of the different frequency divisions. By doing this, the same phase shift is accomplished using half the clocking frequency, and so the power consumption is approximately halved. The second technique uses XOR gating to reduce the effective frequency of the reclocking DFFs that are required to synchronize both F_{drive} and φ_{DAC} to the same unskewed clock. These two signals cannot be clocked by any of the divided frequencies since these have clock-to-q delays on them which are very large. The clock delay of a clock gate is the same as that of an AND-gate, and so it can be used to reduce the clock rate of the synchronizing DFFs. XOR gating then determines if these DFFs must be reclocked by comparing their input to their output.

5.4: Measurements

The design has been realized in a 65 nm technology (see Figure 5.28). The complete sensor area is 2210 µm², of which 60% is consumed by the PPF. As in the previous designs, multiple sensors are placed on a single die. This corresponds well with the thermal management application.

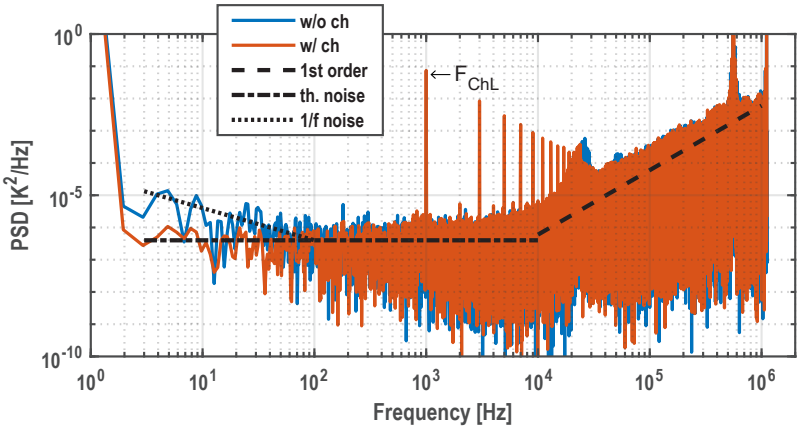

Figure 5.29: PSD of the sensor's bitstream (Hanning window, 1,048,576 samples).

More importantly, it helps simplify the measurement by minimizing the number of oven runs required. In this case, there are 24 samples on a die and four dies have been measured, i.e., a total of 96 samples have been measured.

5.4.1 – Ringing

The bond-wire ringing that is present in previous chips (Section 4.6.1) is also present in these chips. In previous designs, external damping resistors were used to dampen the oscillations in the analog domain. However, activity in the digital domain couples into the analog domain which causes severe residual oscillations. In this design, the digital power consumption is greatly reduced. Furthermore, the on-chip by-pass capacitors have increased in size (from 67 pF to 177 pF), and as a result the effects of the bond-wire oscillations are much less severe.

5.4.2 – Resolution

Figure 5.29 shows the bitstream's PSD under two conditions: with and without *ChL*. In both cases it shows first-order noise shaping. When the low-frequency chopping is enabled, it shows the tones of the low-frequency chopping, which corresponds to a GRO frequency offset of 0.04%. None of the observed samples had a frequency offset worse than 0.5%. Figure 5.30 shows the resolution vs. conversion time of the sensor under the same two conditions. It also compares them to the simulated value (dotted lines). It achieves a 12.8 mK resolution in a 1 ms conversion time.

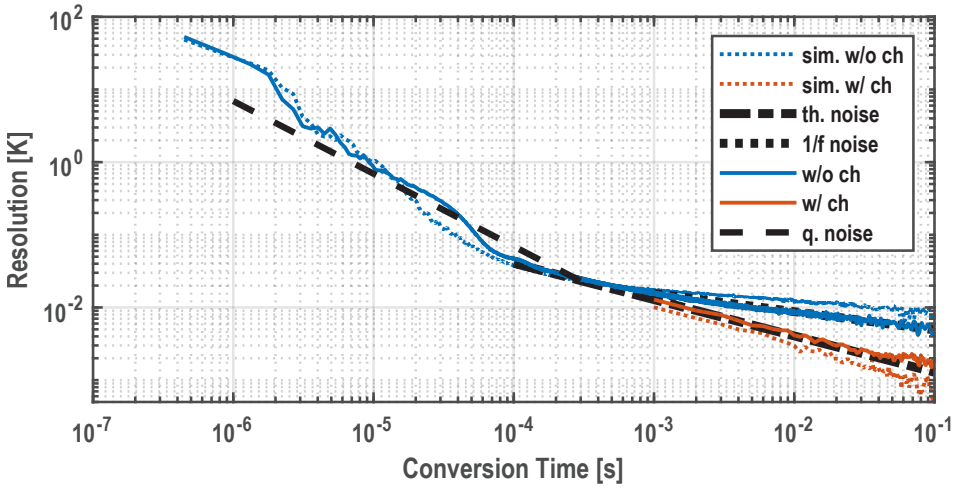

Figure 5.30: Sensor's resolution vs. conversion time.

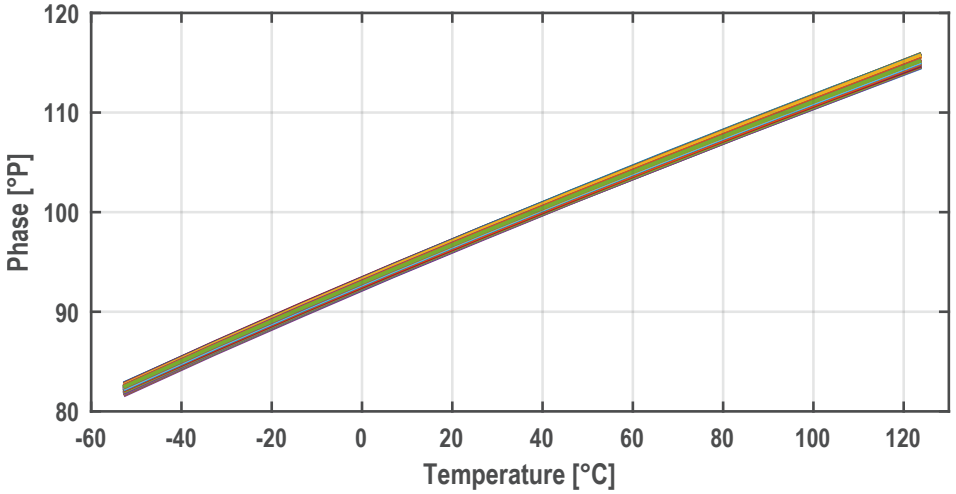

Figure 5.31: Sensor's output phase over temperature.

5.4.3 – Accuracy

Figure 5.31 shows the measured phase over the extended military temperature range (-55°C to 125°C) of the measured samples. Over this range, the samples show a phase range of 33°. The non-linearity that remains after a 1-pt correlated trimming can be explained by the PPF ZC

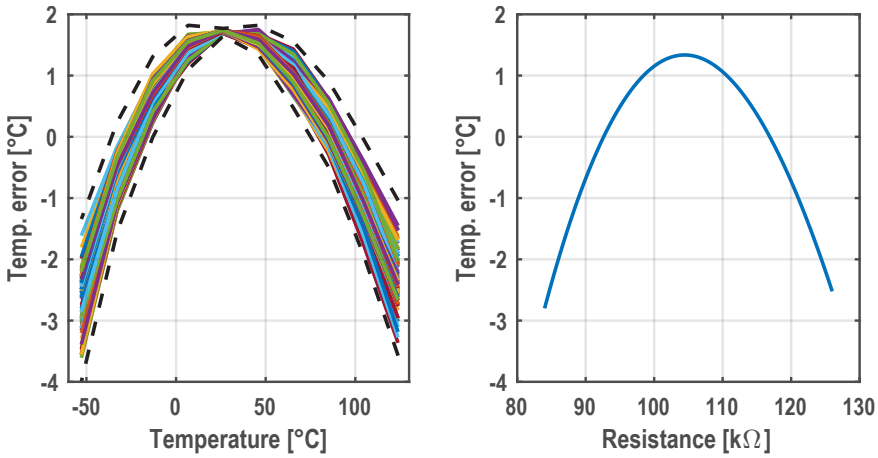

Figure 5.32: Sensor error after a correlated 1-pt trim (left), and simulated non-linearity of the sensor (right).

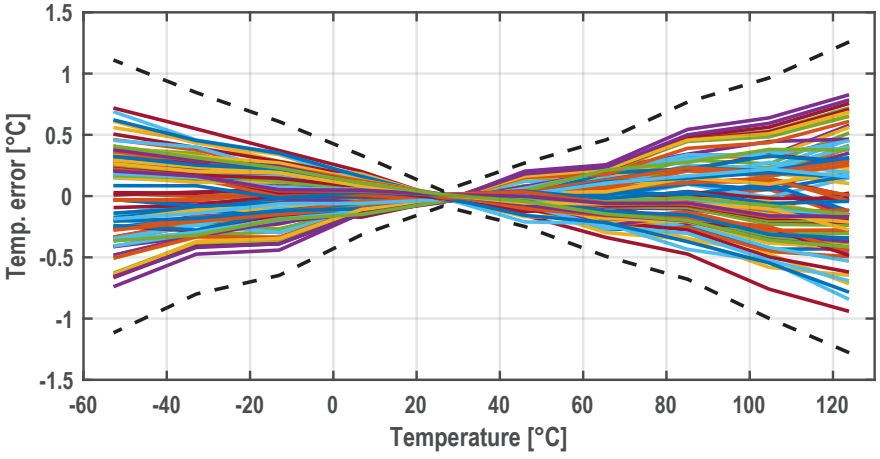

Figure 5.33: Temperature error after a correlated 1-pt trim + 3rd order polynomial non-linearity correction.

non-linearity (see Figure 5.32). The sensor achieves a 3σ error of ±1.3°C after a correlated 1-pt trim and subsequent third-order non-linearity correction (see Figure 5.33). Some of this error is due to residual power-supply ringing, which causes a residual phase error in each measurement, which in turn also reduces the effectiveness of the trimming. This is more clearly illustrated in Figure 5.34, which shows the error after a 2-pt trim and subsequent third-order non-linearity correction. The resulting error is mostly dominated by the effect of the residual ringing.

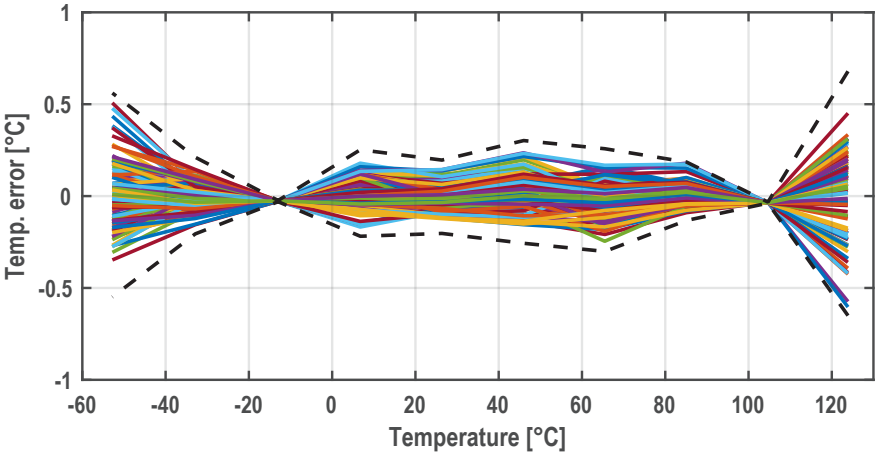

Figure 5.34: Temperature error after a 2-pt trim + 3rd order polynomial non-linearity correction.

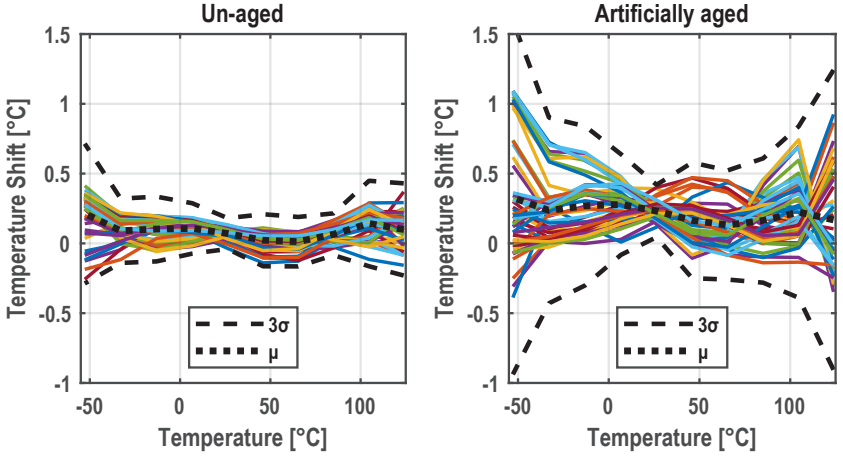

Figure 5.35: Sensor's shift in temperature reading after artificial aging (150°C for one week, right) and un-aged control group (left).

5.4.4 – Aging

The chips were artificially aged in an oven at 150°C for one week. The temperature reading of each sample before aging was subtracted from the temperature reading after aging to determine the shift in the temperature of each sample (see Figure 5.35). The same was also done with a control group that had not been artificially aged between the two measurements. It is not clear what proportion of the spread after aging is due to power-supply ringing, which causes extra error, or to variations in aging. The two

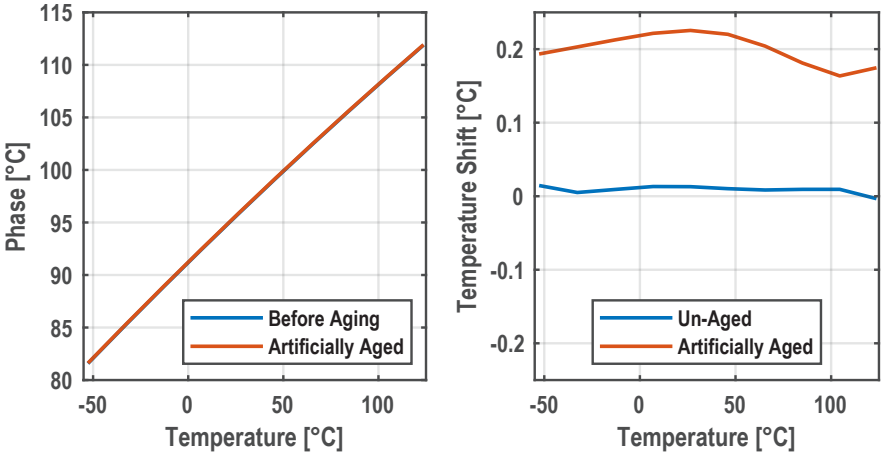

Figure 5.36: Sensor's master curve before and after aging (left), and the temperature error it represents (right).

master curves (before and after aging) and the corresponding temperature shift are shown in Figure 5.36. The shift in the master curve is 0.2°C and remains relatively flat over temperature.

5.5: Conclusion

The goal of this design is to remove the effects of the TQ noise caused by the counter while maintaining the sensor's simplicity. Although some residual TQ noise remains due to the *ChL*, its effect is drastically reduced compared to the original TQ noise. Furthermore, a sinc² window eliminates most of its remaining effect. As a result, this sensor achieves a resolution of 12.8 mK in 1 ms, which is a ~3x improvement over the previous design (see Table 5.1). At the same time, the power consumption is reduced from 134 µW to 28 µW, which is mostly a result of the reduced counting frequency that is facilitated by the reduced TQ noise. More importantly, the design remains a simple and scalable design that uses mostly synthesized digital logic. The only additional circuitry outside of the digital domain is the dual-GRO, but even that consists of inverters. Moreover, even though it requires a bias current, it is insensitive to variations in that bias current. The main downside is the requirement for a static comparator, which has a worse PSS (9.1°C/V) than the dynamic comparator (0.64°C/V) (see Section 4.7.3).

Table 5.1: Performance Summary and Comparison with State-of-the-Art.

Publication	This Work	Chapter 4	ISSCC20 Khashaba [1]	SSCL19 Lee [2]	JSSC17 Sonmez [4]	SSCL19 Eberlein []
Sensor Type	Resistor PDΣΔM	Resistor PDΣΔM	Resistor FLL	Resistor FLL	TD PDΣΔM	Diode SAR
Process	65 nm	65 nm	65 nm	65 nm	40 nm	16 nm
Area [μm²]	2,210	1,930	8,800	5,800	1,650	2,500
Temperature Range (°C)	-55 to 125	-55 to 125	-30 to 90	-50 to 105	-55 to 125	-15 to 105
Inaccuracy [°C]	±1.3 (3σ)	??	±0.72 (p2p)	±1.2 (3σ)	±1.4 (3σ)	+1.5/-2.0 (p2p)
Trimming	1-pt	1-pt	1-pt	1-pt	0-pt	1-pt
Power [μW]	28	134	45	32.5	2,500	18
Supply Voltage [V]	0.9	0.9	1	1	1.05	0.95
Conv. time [ms]	1	1	1	1	1	0.13
Resolution [mK]	12.8 (rms)	40.6 (rms)	1.43 (rms)	2.8 (rms)	360 (rms)	300 (lsb)
PSS [°C/V]	9.1	0.7	2.2	0.22	2.8	1.5

5.6: References

[1] M. Z. Straayer and M. H. Perrott, "A Multi-Path Gated Ring Oscillator TDC With First-Order Noise Shaping," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 4, pp. 1089-1098, 2009.

[2] C. van Vroonhoven, (2015) *Integrated Temperature Sensors based on Heat Diffusion* [Doctoral dissertation, Delft University of Technology].

[3] M. A. Z. Straayer, (2008) *Noise Shaping Techniques for Analog and Time to Digital Converters Using Voltage Crontrolled Oscillators* [Doctoral dissertation, Massachusetts Institute of Technology].

[4] P. Shanthi, "Excess Loop Delay Compensation in Continuous-Time Delta-Sigma Modulators," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 55, no. 11, pp. 1119-1123, 2008.

[5] W. Choi, Y. Lee, S. Kim, S. Lee, J. Jang, J. Chun, K. A. A. Makinwa, and Y. Chae, "A Compact Resistor-Based CMOS Temperature Sensor With an Inaccuracy of 0.12°C (3σ) and a Resolution FoM of 0.43 pJ·K² in 65-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 12, pp. 3356-3367, 2018.

[6] Y. Lee, W. Choi, T. Kim, S. Song, K. A. A. Makinwa, and Y. Chae, "A 5800-μm² Resistor-Based Temperature Sensor With a One-Point Trimmed Inaccuracy of ±1.2 °C (3σ) From −50 °C to 105 °C in 65-nm CMOS," *IEEE Solid-State Circuits Letters*, vol. 2, no. 9, pp. 67-70, 2019.

[7] A. Khashaba, J. Zhu, A. Elmallah, M. Ahmed, and P. K. Hanumolu, "3.2 A 0.088mm² Resistor-Based Temperature Sensor Achieving 92fJ·k² FoM in 65nm CMOS," in *IEEE International Solid-State Circuits Conference*, San Francisco, 2020.

[8] U. Sönmez, F. Sebastiano, and K. A. A. Makinwa, "11.4 1650μm² thermal-diffusivity sensors with inaccuracies down to ±0.75°C in 40nm CMOS," in *IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, 2016.

[9] M. Eberlein and H. Pretl, "A No-Trim, Scaling-Friendly Thermal Sensor in 16nm FinFET Usig Bulk Diodes as Sensing Elements," *IEEE Solid-State Circuits Letters*, vol. 2, no. 9, pp. 63-66, 2019.

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

Chapter 6 | Main Findings and Future Work

In this thesis, a number of compact resistor-based temperature sensors for thermal management have been described. Compared to the state-of-the-art, these sensors reduce the temperature sensor's footprint at a minimal cost in accuracy, improving the area-vs.-accuracy tradeoff. Both low area and a high level of digital integration have been achieved by combining an RC resistor-based temperature sensing front-end with a highly digital and compact version of the Phase-Domain Sigma Delta Modulator (PDΣΔM).

This thesis describes a series of steps to achieve an area-efficient resistor-based temperature sensor. First, the resistor-based readout was miniaturized (Chapter 3). Second, The amount of inaccuracy contributed by the readout circuit has been mitigated by placing the current-controlled oscillator (CCO) outside the loop so that its nonlinearity cannot corrupt the phase of the front end (Chapter 4). Third, the quantization noise impact of the CCO has been mitigated by replacing it with two gated ring oscillators that prevent time-quantization noise from being integrated by the PDΣΔM (Chapter 5).

The main findings, some design considerations, and future work of this thesis are outlined in the following sections.

6.1: Main Findings

The main findings of this work are listed below:

- The silicided poly resistor can be miniaturized without a significant penalty to its spread over temperature. However, its accuracy is worse compared to non-miniaturized versions, and especially its single point trim suffers from miniaturization. This accuracy penalty is likely mainly due to the use of a compact readout, and not solely due to shrinking the resistor area (Sec. 3.3.1, 4.4).

- The nonlinearity of the current-controlled oscillator (CCO) used in the signal path of the highly digital PDΣΔM significantly impacts the sensor's accuracy. In the prior art, the impact of the nonlinearity is insignificant since it was used to read out thermal-diffusivity-based sensors, which have very weak signals (few mVs). In contrast, the current (or voltage) outputs of resistor-based front-ends are very strong signals (100s of mVs) that swing over a larger range of the CCO. Consequently, it has a more significant impact on the accuracy of these types of sensor front-ends (Sec. 3.3.1).

- In high-SNR designs, the time-quantization noise associated with the highly digital PDΣΔM is not sufficiently random. Consequently, the time-quantization noise becomes very tonal, resulting in limit-cycling behavior and large dead zones in the modulator's output. Interestingly, the modulator only shows an indication of this phenomenon at particular frequency ratios between the

oscillator's and PDΣΔM's frequency. Specific frequency ratios will result in tiny dead zones that are barely noticeable (less than a few tens of mK), while other frequency ratios will result in dead zones of multiple °C. Unavoidably, the temperature dependence of the oscillator's frequency will cause it to reside at detrimental frequency ratios at numerous points along the temperature range (Sec. 4.5, 4.6.2).

- The time-quantization noise of the highly digital PDΣΔM can be greatly reduced by locking the oscillator's phase to the counter state. This requires two oscillator/counter pairs resulting in a minor area cost. However, as a result of this quantization noise reduction, the oscillator can run at a much lower frequency, improving the highly digital PDΣΔM's energy efficiency by two orders of magnitude and preventing dead zones due to the limit-cycle behavior (Ch. 5).
- The silicided poly resistor aging seems to be within the thermal management's temperature sensor accuracy budget. In a preliminary test, four chips (each containing 24 sensors) were aged at 150°C for a week showing that the errors due to sensor aging are less than the temperature sensor's initial inaccuracy (Sec. 5.4.4).

6.2: Design Considerations

During the work for this thesis, some insights were gained about the design and optimization of separate blocks. Although they are not main findings, they are still interesting and significant enough to be mentioned:

- When a PDΣΔM is used to detect the phase of a signal that is digital in amplitude (but analog in phase) and when the input phase range is very small (as it is for the front-end in Ch.4), then the polarity of integration (in the PDΣΔM) is fixed for a relatively large part of the period. During these parts of the period, the integrator can be disabled (windowed) without changing the phase detection. Since the integrator is the primary power consumer, this results in lower power. Interestingly, although there are more moments during the PDΣΔM period when time-quantization noise is injected, the total

(linearly summed) time-quantization noise remains the same. Since noise is uncorrelated, the resulting RMS-summed noise is smaller, even though the power is also reduced (Sec. 4.2).

- Optimizing a dynamic (clocked) comparator to accurately detect the phase of an RC-filter has some small but significant differences from designing a comparator that compares the voltage of a quasi-static signal. Since a comparator compares the input with an imperfect impulse response (i.e., one with a finite timing width), it compares the average of the input voltage within that timing window. Usually, to reduce noise and offset, this response/window is purposely smeared out over a considerable amount of time. However, due to the non-linear exponential settling of the RC filter's output and the poor controllability of the time window, this causes an effective phase delay that limits accuracy. Consequently, comparator noise and offset should be sacrificed to achieve accuracy. However, in the highly digital PDΣΔM, this is not an issue since the time-quantization noise of the PDΣΔM dominates (Sec. 4.4.2).

- The poly-phase filter (PPF) is not recommended for thermal management applications. The beyond-the-power-rails swing complicates the readout, and its all-pass filter behavior makes it very susceptible to high-frequency aggressors. One example of such an aggressor is the supply ringing due to the bond wire inductance. The cause of this ringing is the periodic power consumption of the digital logic and the PPF itself (Sec. 4.6.1).

6.3: Future Work

The section below details future work that results from the work for this thesis. To bring some structure, these future work items have been divided into three categories: front-end, readout, and physics.

6.3.1 – Front-End

- The PPF is an all-pass filter that makes the sensor susceptible to high-frequency aggressors. Consequently, future work should use low-pass filters (or band-pass filters) as their sensing front-end. As an additional benefit, this makes the sensor much less susceptible to the ringing from bond-wire inductance, further improving the sensor's accuracy.

- A resistor-based sensor could be combined with a thermal-diffusivity-based temperature sensor to improve the trim-ability of the sensor. Although it would allow for more accurate trimming, the main advantage is that it enables a 0-pt temperature trim, which is required in some thermal management applications. This is because leakage can cause significant self-heating, which means that the internal die temperature does not match the ambient temperature. Because the sensor already contains a PDΣΔM and temperature sensing resistors, it can easily be combined with a heater located at a convenient distance from the front-end resistor. The input signal is then not due to the applied voltage but due to the resistance swing, which in turn is due to the thermal swing caused by the heater. In terms of area, the only additional cost is the area of the heater (which is just a strategically placed switch between the VDD and GND) and a few switches to reconfigure the front end.

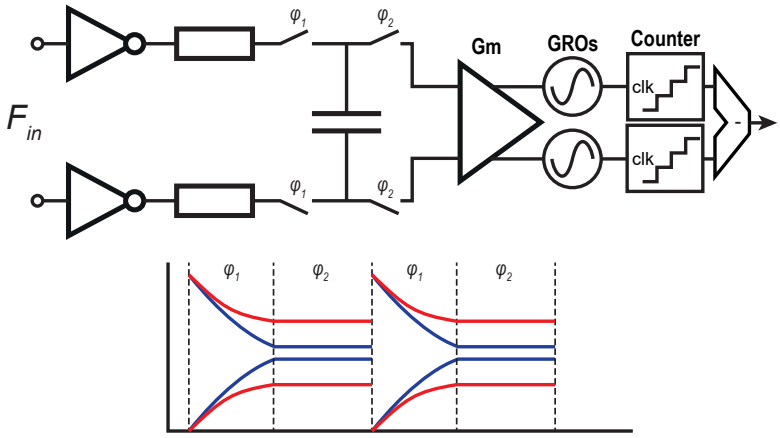

Figure 6.1: Suggested sampled time-constant front-end + 1st order loop-filter.

6.3.2 – Readout

- In the design described in Chapter 5, the continuous-time comparator is the main performance bottleneck on power, noise, accuracy, and the DC PSRR. Alternatively, the output of the filter could be sampled after a fixed delay which could then be used to drive an oscillator/counter pair (in a fully differential setup) similar to [1], see Figure 6.1. In theory, discounting readout power and noise, this readout architecture achieves the best energy FoM among the known resistor-based temperature sensors. In this case, a multi-bit resistor/capacitor/phase-DAC feedback should be used to minimize the inaccuracy due to CCO nonlinearity.

- A second-order loop filter should be used to avoid the limit-cycle behavior of a first-order loop, see Figure 6.2. This minimizes the time-quantization artifacts described in Chapter 4 and reduces the normal ΣΔM quantization noise, which will otherwise be the noise performance bottleneck. It also pairs well with the multi-bit DAC since a digital integrator (as the second integrator) can provide any number of bits required by the DAC.

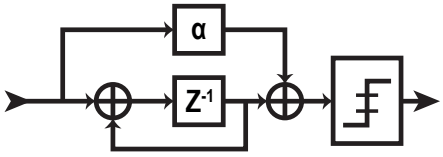

Figure 6.2: Suggested 2nd order loop-filter.

6.3.3 – Physics

- The aging, drift, and mechanical stress sensitivity of the sensor should be studied further. At the time of writing this thesis, the level of aging and drift of these sensors was still not sufficiently understood. Most industrial applications require high confidence in a 10-year accuracy stability. Also, distributing the resistors over two (orthogonal) directions can reduce their overall stress sensitivity.

- In advanced CMOS technologies, the gates are not made from (silicided) poly but are 'metal gates'. The temperature coefficient of these metal gates is much smaller than that of silicided poly resistors and may therefore be unsuitable for temperature sensing. Consequently, an alternative should be found. There are two promising candidates: 1) The metal interconnect should have the best-in-class temperature coefficient accuracy but requires very narrow and long geometries. Therefore, such sensors can only work in technologies with a narrow interconnect size and spacing rules (e.g., [2]). The open question is how much *phonon scattering* will affect the resistance and its temperature coefficient. 2) Alternatively, the NWELL can be used as a temperature-sensing resistor. Although it has a good temperature coefficient, it will have more substrate noise coupling, leakage current into the substrate, higher sensitivity to mechanical stress, and a junction capacitance. Although it has larger size/spacing rules, it also has higher sheet resistance than the silicided poly resistor.

I hope that the combined lists of future work will motivate further research on compact temperature sensors for thermal management. The improvement that this will bring (good PSRR and better efficiency) will allow such sensors to be placed easily and ubiquitously in any SoC, even battery powered ones. This, in turn, will enable better monitoring and therefore better performance of these kind of devices.

6.4: References

[1] A. Khashaba, J. Zhu, A. Elmallah, M. Ahmed, and P. K. Hanumolu, "3.2 A 0.088mm² Resistor-Based Temperature Sensor Achieving 92fJ·k² FoM in 65nm CMOS," in *IEEE International Solid-State Circuits Conference*, San Francisco, 2020.

[2] J. Park, J. Kim, K. Kim, J.-H. Yang, M. Choi, and J. Shin, "A 0.65V 1316μm² Fully Synthesizable Digital Temperature Sensor Using Wire Metal Achieving 0.16nJ·%²-Accuracy FoM in 5nm FinFET CMOS," in *International Solid-State Circuit Conference*, San Fransisco, 2022.

Appendix I | Summary

Advanced Systems-on-Chip (SoCs) exhibit significant self-heating and require thermal management to prevent overheating. Sensors for this application should be not only small so that they can be placed ubiquitously, but they should also achieve moderate accuracy and resolution within 1 ms so that the die temperature can be monitored with approximately 1°C of measurement error. BJTs are the traditional elements that are used to sense temperature, but they require large headroom and are difficult to miniaturize. Thermal diffusivity relies solely on lithography for its accuracy and is easy to miniaturize, but it requires a significant amount of power (a few mWs) to generate the local heat pulses that it uses to measure temperature. MOS-based temperature sensors do not require the headroom that BJT-based sensors do, but they do require a large area to minimize the effect of aging and drift. Resistors achieve superior energy efficiency, moderate accuracy (enough for thermal management), and have no headroom requirements. However, at the beginning of this work, these types of sensors were not yet miniaturized.

Chapter 2 investigates the physical principles of electrical conduction to give an understanding of the different types of resistors that are available and their merits as temperature sensing element. It then investigates the two main ways of digitizing resistance: 1) in the amplitude domain, which requires a resistive reference which limits area efficiency, and 2) in the time domain, which requires a complex impedance and a frequency as a reference, both of which are readily available in SoCs and are very temperature-stable. The state-of-the-art uses three different types of architectures to combine resistors and capacitors to digitize resistance, each having their own merit: 1) a low-pass filter, 2) a Wien-Bridge band-pass filter, and 3) a poly-phase all-pass filter. Digitization can be done in two different ways: 1) by employing the filter in a frequency-locked loop, which generates a temperature-dependent output that is digitized by a subsequent frequency counter, or 2) directly digitizing its phase with a phase-domain sigma-delta modulator (PDΣΔM). The latter comes in a highly digital version that can be made very small, but exhibits counter wrap-around and suffers from time-quantization noise.

Chapter 3 describes the design of the first attempt at miniaturizing a resistor-based temperature sensor. It uses a single-ended Wien-Bridge, which is enabled by the Wien-Bridge's pass-band filter that prevents DC power-supply variations. A highly digital PDΣΔM is used to digitize the Wien-Bridge phase response. Careful design of the counter prevents wrap-around errors in the integrator while careful design of the CCO's center frequency and its frequency swing minimizes the time-quantization noise. Although time-quantization noise requires large frequency swing in the CCO to minimize its noise contribution, the CCO's non-linear current-to-frequency transfer along with the large swing causes significant inaccuracy in the readout. As a result, after a 1-point trim it achieves a 3σ accuracy of 2.7°C from -35°C to 125°C and 1.2°C after a 2-point trim.

In Chapter 4, to improve the accuracy, the CCO is placed outside of the PDΣΔM loop. As an additional benefit, this readout does not suffer from cosine non-linearity which enables its use in a multi-bit system. Additionally, the counter integrator is windowed to save power and reduce time-quantization noise. To minimize the supply sensitivity and improve accuracy, a dynamic comparator is used. To enable these improvements, careful design of its delay, its charge kick-back, and the input over-voltage handling must be done. Due to the high SNR input to the highly digital PDΣΔM, its time quantization noise causes timing artifacts in the modulator, which in turn causes large dead-zones in its phase readout and high tonality in its output bitstream. These effects are minimized by picking good frequency ratios between the oscillator and the PDΣΔM sampling frequency.

The design in Chapter 5 aims to solve the time quantization noise of the PDΣΔM. This relaxes the oscillator speed requirements and hence its power consumption, it improves the sensor's resolution as its noise was dominated by the time quantization noise, and it prevents the timing artifacts that were seen in Chapter 4. To accomplish this, it requires two oscillator and counter pairs so that the oscillator state is locked to the counter state. A dual-gated ring oscillator is used to prevent the need for a current-controlled oscillator. The dual-GRO is chopped to prevent frequency drift and offset from impacting the accuracy. Careful design of the bias noise, GRO frequency, parasitic capacitance of the GRO nodes, and leakage of the GRO switches minimizes their noise contribution. Careful design of the digital blocks (the counter, summation node, and phase-DAC) minimizes their power

consumption. Consequently, the sensor consumes 28 µW (down from 134 µW in the previous chapter) and achieves a 12.8 mK resolution within a 1 ms conversion time (down from 40.6 mK in the previous chapter). It achieves an accuracy of 1.3°C (3σ) from -55°C to 125°C after a 1-point trim, while only consuming an area of 2210 µm² in a 65-nm technology.

Appendix II | Samenvatting

Geavanceerde chips zoals processoren vertonen significante hoevelheid zelf-verwarming die voorkomen moet worden met behulp van temperatuursensoren. Deze sensoren moeten compact zijn zodat ze arbitrair geplaatst kunnen worden. Tegelijkertijd moeten ze ook gematigd accuraat en precies zijn zodat de totale meetfout in een 1 ms tijdseinheid niet groter is dan 1°C. Bi-polaire transistoren (BJTs) zijn het meest gebruikte element om een temperatuursensor te maken, maar ze hebben een groot potentiaalverschil nodig en zijn moeilijk om compact te maken. Op Thermische difusie gebaseerde temperatuursensoren vertrouwen op de accuraatheid van litografie voor hun eigen acuraatheid and zijn zodoende makkelijk the verkleinen, maar hebben een redelijke hoeveelheid vermogen nodig (enkele mW) om de lokale hittepulsen te genereren die gebruikt worden om de temperatuur te meten. MOSFET gebaseerde temperatuursensoren hebben niet de hoofdruimte nodig die BJT's nodig hebben, maar hebben veel oppervlak nodig om veroudering en verschuiving binnen de perken te houden. Weerstands-gebaseerde temperatuursensoren behalen superieure energie-efficientie, gematigde accuraatheid (voldoende voor thermisch beheer) en hebben geen hoofdruimte vereiste. Echter, aan het begin van dit werk waren deze type sensoren nog niet verkleint.

Hoofdstuk 2 onderzoekt de fysische beginselen van electrische conductie om een beter begrip te krijgen van de verschillende weerstanden die beschikbaar zijn en hun verdienste als temperatuursensorelement. Daarna analyseert het de twee voornaamste methodes om weestanden te digitaliseren: 1) in het amplitude domein waarbij een referentie weerstand nodig is en 2) in het frequentie domein waarbij een complexe impedantie en een frequentie als referentie nodig zijn, waarvan de laatste twee vrijelijk aanwezig zijn in geavanceerde chips en ook nog eens temperatuurs-onafhankelijk zijn. In literatuur worden drie verschillende architecturen gebruikt om weerstanden en capaciteiten te combineren om te gebruiken als temperatuursensor, elk met zijn eigen verdienste: 1) het laagdoorlaat-filter, 2) een Wien-Brug band-doorlaat filter en 3) een meer-fase (alles-doorlaat) filter. Digitalisatie van de weerstand kan gedaan worden op twee verschillende manieren: 1) door het gebruik van een frequentie

vergrendelde lus die een weerstandsafhankelijke frequentie genereerd die vervolgens gedigitaliseerd word door een frequentieteller. 2) door gebruik van een fase-domein sigma-delta modulator (PDΣΔM) die direct de fase digitaliseerd. De laatste bestaat ook in een hoogst digitale versie die erg compact is, maar wel teller omwikkelingen vertoont en leidt onder tijds-kwantisatie ruis.

Hoofdstuk 3 omschrijft het ontwerp van een eerste poging tot het verkleinen van een weerstands-gebaseerde temperatuursensor. Het gebruikt een enkelzijdige Wien-Brug, wat mogelijk gemaakt wordt door het band-doorlaat filter dat voorkomt dat DC voedingsvariaties de meeting beïnvloeden. Een hoogst digitale versie van de fase-domein sigma-delta modulator is gebruikt om de Wien-Brug faserespons te digitaliseren. Zorgvuldig ontwerp van de teller voorkomt teller omwentelingen en zorgvuldig ontwerp van de centrale frequentie van de stroom gecontrolleerde oscillator en zijn frequentiezwaai minimaliseert de tijds-kwantisatieruis. Alhoewel voor het verminderen van de tijds-kwantiesatieruis een groote frequentiezewaai nodig is, vereist de niet-lineariteit van de stroom gecontrolleerde oscillator juist dat er een kleine frequentiezwaai is om zijn gevolgen op de acuraatheid te verminderen. Als gevolg van deze acuraatheid behaald de sensor een 3σ temperatuurs acuraatheid van 2.7° over een temperatuursbereik van -35°C tot 125°C na een 1-punts kalibratie en een acuraatheid van 1.2°C na een 2-punts kalibratie.

In hoofdstuk 4 is the stroom gecontrolleerde oscillator buiten de PDΣΔM lus geplaatst om de accuraatheid te verbeteren. Als bijkomstig voordeel leidt deze aanpassing niet aan cosine niet-lineariteit wat het mogelijk maakt om gebruikt te worden in een multi-bits system. Daarbovenop wordt de teller/integreerder periodiek geactiveerd wat energie bespaart en tijds-kwantisatie ruis vermindert. Om de voedings voltage gevoeligheid te verminderen en accuraatheid te verbeteren, wort een dynamische comparator gebruikt. Om van deze verbeteringen gebruik te kunnen maken, is zorgvuldig ontwerp van zijn vertraging, ladingsterugslag, en de overvoltage bescherming nodig. Vanwege de hoge ruis-to-signaal verhouding van het ingangssignaal veroorzaakt de tijdskwantisatieruis van de PDΣΔM voor tijdsartefacten die dode zones en tonaliteit in de bitstream veroorzaakt. Het effect van deze tijdsartefaction kan vermindert worden door het uitkiezen van optimale frequentieratios tussen de oscillator en PDΣΔM bemonsteringsfrequentie.

Het ontwerp in hoofdstuk 5 doelt zich om de tijdskwantisatie van de PDΣΔM te verhelpen, wat de snelheidsvereiste van de oscillatorfrequentie vermindert en daarmee ook de sensor's ruis en vermogensverbruik vermindert. Bijkomend voordeel is dat het ook de tijdskwantisatie effecten van hoofdstuk 4 voorkomt. Om dit voorelkaar te krijgen zijn twee oscillator en teller paren nodig zodat de toestand van beide oscillator en teller paren aan elkaar gebonden zijn. Een paar aan geactiveerde ring oscillatoren wordt gebruikt om te voorkomen dat er twee oscillatoren naast elkaar aan staan. Deze worden periodiek omgewisseld om laag frequent ruis en asymetrische fouten te voorkomen. Zorgvuldig ontwerp van de referentiestroomruis, oscillator frequentie, parasitaire capaciteiten van de oscillator, en de lekkagestroom van de oscillatorschakelaren verminderen hun ruiscontributie. Zorgvuldig ontwerp van de digitale blokken (de teller, sommatieblok, en phase digitaal-naar-analoog omzetter) verminderen het vermogensverbruik. Als gevolg verbruikt de sensor 28 µW (verminderd van 134 µW in het vorige hoofdstuk) en bereikt een resoutie van 12.8 mK in een 1 ms meettijd (verminderd van 40.6 mK in het vorige hoofdstuk). De sensor behaalt een accuraatheit van 1.3°C (3σ) tussen -55°C en 125°C na een 1-punts trim, terwijl het een oppervlatke verbruikt can maar 2210 µm² in een 65-nm technologie.

Acknowledgements

No good work worthy of praise is made alone and in isolation. It requires debate and collaboration, not to mention reflection, introspection, and distraction. This chapter is my chance to acknowledge those that helped me accomplish my goals.

First and foremost, I want to acknowledge my promoter Kofi A. A. Makinwa. He has supported me, believed in me, and kept me on track even in the darkest of times. His unwavering dedication to pushing the boundaries of the State-of-the-Art, his ingenuity and breadth of technical knowledge, and especially his inexhaustible and unwavering attention that he gives to any person, irrespective of their technical competence, has always been an inspiration to me and will continue to do so. Although he might be tough and demanding, I have never found him to ask for more than he is willing to give. Moreover, he has always had my back and supported me in challenges that I faced even, and especially, outside of academia. Thank you, Kofi, for showing and teaching me how a truly accomplished IC Designer works and communicates with those around him.

None of the work that I, or anyone at the Electronic Instrumentation Laboratory, does is possible without the support of the staff. Thank you, Zu-Yao Chang, Lukasz Pakula, Joyce Siemers, Ron van Puffelen, Atef Akhnoukh, and Pieter Frehe. Without your support no servers would be present and maintained, no PDK access would exist, no problems would be solved, no measurements would be possible, no group meetings could be planned, no building access would be available, and no work would be possible. You are the ones that do the work that is most invisible, but most important for all that EI does.

I would also like to give a special thanks to my colleagues for all the discussions and distractions that they provided. Over the years there have been too many to individually mention, but I would like to specifically mention Uğur Sönmez, who was my supervisor during my master's thesis, and who has always been available for a technical discussion. I would also like to thank Burak Gönen, Çağrı Gürleyük, and Lorenzo Pedalà for their support and friendship.

Last, I would like to thank my family, without whom none of this would be possible. They enabled me to pursue the academic degree from which most of my current life followed. Even though they never fully understood what exactly it is that I do (despite their repeated asking), they always stood behind me. Their unwavering support was never based on success but solely based on whether I gave it my best and on my happiness. Thank you for your support without which none of this would have been possible.

List of Publications

Journal Papers

[1] W. Choi, J. Angevare, I. Park, K. A. A. Makinwa, and Y. Chae, "A 0.9-V 28-MHz Highly Digital CMOS Dual-RC Frequency Reference With ±200 ppm Inaccuracy From -40 °C to 85 °C," in *IEEE Journal of Solid-State Circuits*, vol. 57, no. 8, pp. 2418-2428, Aug. 2022.

[2] T. Someya, V. van Hoek, J. Angevare, S. Pan and K. Makinwa, "A 210 nW NPN-Based Temperature Sensor With an Inaccuracy of ±0.15 °C (3σ) from -15 °C to 85 °C Utilizing Dual-Mode Frontend," in *IEEE Solid-State Circuits Letters*, vol. 5, pp. 272-275, 2022.

[3] S. Pan, J. A. Angevare, and K. A. A. Makinwa, "A Self-Calibrated Hybrid Thermal-Diffusivity/Resistor-Based Temperature Sensor," in *IEEE Journal of Solid-State Circuits*, vol. 56, no. 12, pp. 3551-3559, Dec. 2021.

[4] J. A. Angevare and K. A. A. Makinwa, "A 6800-μm² Resistor-Based Temperature Sensor With ±0.35°C (3σ) Inaccuracy in 180-nm CMOS," in *IEEE Journal of Solid-State Circuits*, vol. 54, pp. 2649-2657, Oct. 2019.

Conference Papers

[1] T. Someya, V. van Hoek, J. Angevare, S. Pan, and K. A. A. 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, pp. 120-121, 2022.

[2] J. A. Angevare, Y. Chae, and K. A. A. Makinwa, "5.3 A Highly Digital 2210μm² Resistor-Based Temperature Sensor with a 1-Point Trimmed Inaccuracy of ±1.3°C (3σ) from -55°C to 125°C in 65nm CMOS," *2021 IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, pp. 76-78, 2021.

[3] W. Choi, J. A. Angevare, I. Park, K. A. A. Makinwa, and Y.Chae, "31.2 A 0.9V 28MHz Dual-RC Frequency Reference with 5pJ/Cycle and ±200 ppm Inaccuracy from -40°C to 85°C," 2021 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, pp. 434-436, 2021.

[4] S. Pan, J. A. Angevare, and K. A. A. Makinwa, "5.4 A Hybrid Thermal-Diffusivity/Resistor-Based Temperature Sensor with a Self-Calibrated Inaccuracy of ±0.25°C (3σ) from -55°C to 125°C," *2021 IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, pp. 78-80, 2021.

[5] J. Angevare and K. A. A. Makinwa, "A 6800-μm² Resistor-Based Temperature Sensor in 180-nm CMOS," *2018 IEEE Asian Solid-State Circuits Conference (A-SSCC)*, Tainan, Taiwan, pp. 43-46, 2018.

[6] J. Angevare, L. Pedalà, U. Sönmez, F. Sebastiano, and K. A. A. Makinwa, "A 2800-μm² Thermal-Diffusivity Temperature Sensor With VCO-Based Readout in 160-nm CMOS," 2015 Asian Solid-State Circuits Conference (A-SSCC), Xiamen, China, 2015.

[7] K. Blutman, J. Angevare, A. Zjajo, and N. van der Meijs, "A 0.1pJ Freeze Vernier time-to-digital converter in 65nm CMOS," *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, Melbourne, VIC, Australia, 2014.

[8] J. Angevare, K. Blutman, A. Zjajo, and N. van der Meijs, "A CMOS 0.23pJ Freeze Vernier Time-to-Digital Converter," *2013 NORCHIP*, Vilnius, Lithuania, pp. 1-4, 2013.

About the Author

Jan (Johannes Adrianus) Angevare obtained his B.Sc. and M.Sc. (Dutch: Ir.) degrees in electronics from Delft University of Technology, Delft, the Netherlands, in 2012 and 2015, respectively. From 2016 to 2020, he pursued his Ph.D. at the Electronic Instrumentation Laboratory at Delft University of Technology.

He Joined SiTime in 2021 and has subsequently been working as an analog/mixed-signal IC designer. His current research interests include sensor interfaces, precision analog circuits, data converters, oscillators, and PLLs.