A Single-Trim Frequency Reference System with 0.7 ppm/°C from −63 to 165 °C Consuming 210 µW at 70 MHz

Alexander S. Delke*, Student Member, IEEE, Thomas J. Hoen*, Student Member, IEEE
Anne-Johan Annema, Member, IEEE, Yanyu Jin, Jos Verlinden, Bram Nauta, Fellow, IEEE

Abstract—This paper presents a frequency reference system that combines high frequency accuracy and low power consumption using a single-point temperature trim and batch calibration. The system is intended as a low-cost fully integrated crystal oscillator replacement. In this system, the oscillation frequency of a power-efficient, but Process, Voltage, Temperature (PVT) and Lifetime (L) sensitive current-controlled ring oscillator (CCO) is periodically (re)calibrated by the well-behaved frequency stability of an untuned LC-based Colpitts oscillator (LCO), which is optimized for stability over PVTL. During the single-point room temperature factory trim, the frequency of the LCO is determined and the result is digitally stored. An on-chip calibration engine tunes the ring oscillator to the target frequency based on the LCO frequency, temperature sensor information and digitally stored trimming information, thus effectively improving the frequency stability of the ring oscillator. The relatively high-power LCO is heavily duty-cycled to minimize the overall power consumption. A prototype fabricated in a 0.13-µm high-voltage (HV) CMOS SOI process and assembled in a plastic package demonstrates an inaccuracy lower than ±93 ppm over a temperature range from −63 to 165 °C across 18 samples. The presented frequency reference system, including on-chip voltage regulators and a temperature sensor, occupies a chip area of 0.69 mm² and consumes about 64 µA from a single 3.3 V supply. The frequency error due to supply variation is roughly 92 ppm/V. The mean frequency shift due to aging, measured before and after a six-day storage bake at 175 °C, is only 52 ppm.

Index Terms—Aging, batch calibration, Colpitts oscillator, current-controlled oscillator, frequency reference, LC oscillator, low-power, single-trim, temperature stability, trimming.

I. INTRODUCTION

FREQUENCY references are essential building blocks for many electronic systems for timing and synchronization purposes. The accuracy of frequency references must be high enough to allow other subsystems, e.g., transmitters or analog-to-digital converters (ADCs), to reach specific performance levels. The reference frequency for e.g. SS USB or 10/100/1000 Ethernet must be within ±300 ppm and ±100 ppm margin over PVT variations and lifetime (PVTL) respectively [1]. In general, higher accuracy levels allow for better system performance. To achieve demanding accuracy levels, it is common in the industry to use crystal oscillators (XOs), which have bulky external resonators that cannot be integrated in a CMOS die. These references, however, achieve frequency stabilities better than ±100 ppm over a large temperature range without compensation [2]. Although the size of quartz crystals has shrunk over the past decades, the fabrication method and the need for a vacuum package makes integration challenging [2]. Alternatively, (electrostatic) micro-electro-mechanical systems (MEMS)-based oscillators can achieve a higher level of integration with similar frequency accuracy by compensating for the inherent temperature drift (−20 to −30 ppm/°C) and trimming. These MEMS-based oscillators have gained more market share over recent years; however, the mechanical resonator requires a vacuum cavity, which adds processing steps and costs [3], [4]. Bulk acoustic wave resonator (BAW)-based oscillators (piezoelectric MEMS) with a dual-Bragg bulk acoustic resonator (DBAR) operate without a cavity and can be packaged in relatively low-cost non-hermetic plastic packages [5]. These resonators are stacked as a system-in-package (SiP) solution on top of the CMOS die that contains the active circuitry. This technique increases the level of integration, but the resonator cannot (yet) be integrated with a standard single-die CMOS process flow. Fully integrated frequency references in a standard CMOS process are thermal-diffusivity (TD)-based [6]–[8], RC-based [9]–[17] or LC-based [11], [18]–[24], which differ in power consumption and are all to a greater or lesser extent subject to PVT-variations and degradation over lifetime resulting in inaccurate and long-term instability. Section II extends further on previously reported integrated frequency references.

This paper presents a fully integrated frequency reference system combining the high frequency stability of an LC-based oscillator (LCO) and the low power consumption of an RC-based oscillator, which has similarities with the work in [25]–[27] without relying on an external BAW-based [25], MEMS-based [26], or crystal-based resonator [27]. The design aims to achieve long-term stability over PVTL at a reference frequency of 70 MHz and thus to replace an XO with its accompanying PLL for future/proprietary IoT systems. Its frequency is flexible and can be changed easily via a look-up table (LUT). We demonstrate a frequency inaccuracy better than ±93 ppm from −63 to 165 °C, consuming about 64 µA from a single 3.3 V.
supply with only a single-temperature factory trim (1T-trim) and batch calibration. The mean frequency shift due to aging, measured before and after a six-day storage bake at 175 °C, is within 52 ppm.

The paper is organized as follows: previously reported integrated frequency references are discussed in Section II. Section III introduces the general concept of our frequency reference architecture. Our specific implementation of the system and its key building blocks are discussed in Section IV. Section V presents the measurement results, and Section VI summarizes the main findings and system performance.

II. INTEGRATED FREQUENCY REFERENCES

The primary goal of frequency references is to provide an accurate and stable clocking signal over production- and operational variation as well as lifetime. Figure 1 shows the residual temperature coefficient (TC_R in ppm/°C) after trimming as a function of the power efficiency (in mW/MHz^2) for some fully integrated frequency references. Also included is the temperature range over which the reported TC_R is achieved. As depicted, most highly stable references use batch calibration, which generally improves the residual temperature coefficient. Aging effects are not plotted, as they are generally not reported. This (incomplete) benchmark also doesn’t take the phase noise of the reference frequency into account.

RC-based references typically have the lowest power consumption (due to, in general, a relatively low frequency) but have a residual TC_R in the order of 1 to 10 ppm/°C after a 2T-trim due to complex thermal behaviour of the RC time constant. The oscillation frequency of RC-based references is regardless of the architecture, e.g., open-loop relaxation oscillator [9], [10] or with a frequency-locked-loop (FLL) [11]–[17], inversely proportional to the resistance and capacitance of integrated doped (poly) silicon resistors and capacitors. The resistors are typically the bottleneck for accuracy as these heavily suffer from PVT variation [16], [29], [30]. To minimize temperature dependency, so-called zero-TC (ZTC) resistors may be used that are composed of multiple materials with complementary TC, yielding overall low TC_R [16]. Their eventual (small) TC_R depends on process-sensitive sheet resistances. Typically a multi-temperature trim [16] is required for accuracies better than 1000 ppm over PVT with large (e.g., automotive) temperature ranges. Doped (poly) silicon resistors drift over lifetime due to, e.g., hydrogen release and doping (de-)activation [29], [30]. After a few hours at a high temperature (>150 °C), the resistor value may already drift by 1000 ppm [29], making these references impractical for applications that require stabilities better than ±100 ppm. Resistors (and other front-end components) are also known to have significant (package-)stress dependency [15], [31], [32] that reduce frequency stability, especially in plastic packages.

Many LC-based [19]–[22] and TD-based [6]–[8] references achieve similar or better accuracy using a more cost-effective 1T-trim. This is thanks to their inherently more robust temperature behaviour. TD-based references are based on the thermal diffusivity of bulk silicon and achieve about ±1000 ppm across the military temperature range using a 1T-trim [7]. The frequency accuracy is reported to be mainly limited by the internal temperature sensor required for the correction loop, which causes roughly 500 ppm frequency inaccuracy for 0.1 °C sensing error [6].

The oscillation frequency of LC-based references is primarily defined by their inductance and capacitance, both of which can be implemented in the metal-backend and are independent of doping and semiconductor effects. LC-based references are also less dependent on mechanical (package-)stress compared to RC-based references [15], [19], [31], [33]. In [19], it has been shown that LCOs have a well-defined temperature dependency of roughly 10 to 100 ppm/°C before temperature compensation and have very low degradation over lifetime [18], [19]. The temperature dependency can be compensated with e.g. variable capacitors [20] or with a fractional divider [18], [19]. Both LC-based and TD-based references have a substantial power consumption (>1 mW), with the former to sustain high-frequency oscillation of the lossy LC-tank, and the latter to create a sufficiently large thermal wave [7].

III. PROPOSED SYSTEM ARCHITECTURE

The approach of the presented architecture is to periodically (re)calibrate a relatively PVT-sensitive low power oscillator with a second oscillator that is stable across PVT-variation after the 1T-trim, but that consumes more power. The system, therefore, acts as an FLL which periodically conveys its stability across PVT to the low power oscillator. The relatively high-power oscillator is turned on only for (re)calibrations. Figure 2 shows a simplified block diagram of the architecture conceptually consisting of four main blocks. The first
block is the tunable and always-on low-power oscillator that generates the output frequency $f_{\text{ext}}$. The second block is a frequency sampler\(^2\) that effectively implements an accurate (1T-trimmed) on-chip frequency reference to calibrate the low-power oscillator periodically. The third block is a calibration engine that duty-cycles the frequency sampler and controls the tuning circuitry of the low-power oscillator. The calibration engine is controlled by the fourth block, containing one or more sensor(s); we use only a temperature sensor.

The system conceptually operates as follows; during factory trim at room temperature (RT), an external frequency ($f_{\text{ext}}$) is supplied. The counter then counts the periods of the stable oscillator frequency $f_{\text{st}}$ within a period of $f_{\text{ext}}$ and digitally stores the (fundamentally floored\(^3\)) counter output $\lfloor f_{\text{st}} / f_{\text{ext}} \rfloor$ in memory. Provided that $f_{\text{st}}$ is stable over PVTL or that $f_{\text{st}}$ has a well described dependency on PVTL with accompanying sensors, this trimming approach is equivalent to sampling and holding $f_{\text{ext}}$, hence the name frequency sampling. An advantage of this technique is that it does not require tuning of the accurate oscillator, which makes the oscillator robust over PVTL and makes trimming time-efficient. After the frequency sampler is trimmed, the low power oscillator ($f_{\text{cal}}$) can be sampled (characterized) similarly. The difference between the counter output $\lfloor f_{\text{st}} / f_{\text{ext}} \rfloor$ and the stored value $\lfloor f_{\text{st}} / f_{\text{ext}} \rfloor$ is a measure for the frequency difference between $f_{\text{st}}$ and $f_{\text{ext}}$. By (periodically) sampling and comparing, $f_{\text{st}}$ can be calibrated to be equal to $f_{\text{ext}}$ or to a scaled version thereof.

### A. Duty cycling

The variations over time of $f_{\text{st}}$, as generated by the low-power oscillator, are dominantly temperature dependent, with a relative temperature coefficient (TC) $\text{TC}_{\text{st}}$ (in ppm/°C). The system in Figure 2 could be configured to continuously recalibrate the low-power oscillator to minimize temperature changes on $f_{\text{st}}$: from a system perspective this approach is senseless, but it serves as starting point for the derivation that follows. Defining a relative frequency accuracy constraint $\epsilon_{\text{cal}}$ (in ppm) for $f_{\text{cal}}$ as $|\Delta f_{\text{cal}} / f_{\text{cal}}| \leq \epsilon_{\text{cal}}$ and denoting the time-duration for a (re)calibration as $t_{\text{cal}}$, the maximum rate of temperature change that can be handled by the system, to stay inside accuracy margins is about $\frac{\Delta f}{\Delta t}_{\text{max}} \approx \frac{\text{TC}_{\text{st}} \cdot \epsilon_{\text{cal}}}{t_{\text{cal}}}$.

For the system presented in this paper, see sections IV-B and V. $\epsilon_{\text{cal}} = 100$ ppm, $\text{TC}_{\text{cal}} \approx 242$ ppm/°C and $t_{\text{cal}} \approx 2.6$ ms. This results in a maximum trackable temperature gradient of about 160 °C/s. Tracking a lower temperature gradient $\frac{\Delta f}{\Delta t}$ enables duty-cycling the recalibration of the low-power oscillator at a duty cycle

$$D = \frac{\frac{\Delta f}{\Delta t}_{\text{actual}}}{\frac{\Delta f}{\Delta t}_{\text{max}}} = \frac{\text{TC}_{\text{st}} \cdot t_{\text{cal}}}{\epsilon_{\text{cal}}}$$

(ignoring the loop resolution and noise) while staying within the $\epsilon_{\text{cal}}$ constraint. The average power consumption of the entire system $P_{\text{total,avg}}$ is then

$$P_{\text{total,avg}} = D \cdot P_{\text{cal}} + P_{\text{on}}$$

$$= \frac{\frac{\Delta f}{\Delta t}_{\text{actual}}}{\frac{\Delta f}{\Delta t}_{\text{max}}} \cdot \text{TC}_{\text{st}} \cdot t_{\text{cal}} \cdot P_{\text{cal}} + P_{\text{on}}$$

(2)

where $P_{\text{cal}}$ is the combined power consumption of the frequency sampler and calibration engine and $P_{\text{on}}$ is the power consumption of the low-power oscillator and sensor(s) to activate the (re)calibration.

Figure 3 shows the power consumption of the implemented system and the contribution of its sub-blocks (with $P_{\text{cal}} = 15.7$ mW and $P_{\text{on}} = 0.21$ mW, see Section V-D) versus time spent in low-power mode. The time between (re)calibrations can be long for small variations in temperature, see (1). With a (re)calibration every 1.9 s, the power spent for the (re)calibration is already about a magnitude lower than that of the always-on circuitry.

One might opt for a fixed duty-cycle (re)calibration loop, which is a simple solution that can drop the power consumption easily by an order of magnitude or more. However, this solution can lead to many unnecessary (re)calibrations at the cost of power consumption as well as (re)calibrations only long after $f_{\text{st}}$ got outside the accuracy constraint. To circumvent this, one or more sensors can be employed that trigger (re)calibrations when deemed necessary, e.g. when the temperature has changed more than $\epsilon_{f_{\text{cal}}} \approx 0.4$ °C (for our implementation) with respect to the previous (re)calibration.

A further decrease in power consumption could be accomplished by storing and interpolating between e.g. the most recent or most frequently used calibration values for $f_{\text{cal}}$, which would reduce the need to start the relatively power-hungry frequency sampler while still being able to maintain high accuracy. More complicated systems could employ machine learning [27].

### IV. Circuit Implementation

Our specific implementation of the generic system as discussed in Section III is explained in this section. The presented frequency reference system (see section III) is implemented in a 0.13 µm HV CMOS SOI process\(^4\) and has a single 3.3 V supply. For this prototype, we aim at an output frequency of

\(^2\)The name frequency sampler will be made clear below.

\(^3\)The floor function is indicated by the \(\lfloor \cdot \rfloor\)-brackets.

\(^4\)To the best of the authors knowledge, there is no specific benefit of using an HV SOI process for this system.
70 MHz with a 1T-trimmed accuracy better than ±100 ppm across the automotive temperature range and lifetime. In our implementation, a current-controlled ring oscillator (CCO) is used as a relatively low-power oscillator. Its supply current is derived from a bandgap voltage over a resistor, making it an RC-type oscillator. This CCO is occasionally (re)calibrated by the well-behaved frequency of an untuned LC-based Colpitts oscillator. Due to the choice for an LCO instead of a TD-based oscillator for the highly stable reference frequency, the temperature resolution requirements of the accompanying temperature sensor can be relaxed by about a factor of 100. The relatively high-power Colpitts oscillator is heavily duty-cycled to minimize the overall power consumption.

The block schematic of our system is shown in Figure 4 and consists of a digital core of the calibration engine, the LCO, the CCO, a temperature sensor and voltage regulators. Additional digital circuitry for the temperature sensor and a LUT for the calibration engine are implemented off-chip for evaluation reasons. The remaining parts of this section discuss the different blocks and their workings in more detail.

A. Frequency sampler

The core of the system is the frequency sampler that enables periodic calibration of the CCO that provides the clock output \( f_{\text{ref}} \). During production trimming, an external frequency \( f_{\text{ext}} = 1 \text{ MHz} \) divided by \( N_{\text{ext}} = 1000 \) starts/stops the counter that counts edges of \( f_{\text{st}} \). After one period of \( f_{\text{ext}}/N_{\text{ext}} \), the counter value \( M_{\text{counter}} \) is stored as \( M_{\text{trim}} \):

\[
M_{\text{trim}} = \left[ f_{\text{st}} \left( P_{\text{sample}} \cdot V_{\text{trim}} \cdot t_{\text{trim}} \cdot t_{\text{trim}} \right) \cdot N_{\text{ext}} \right] \quad \text{f}_{\text{ext}}
\]

where \( M_{\text{trim}} \) is a digital representation for the process spreaded \( f_{\text{st}} \) (about 250 MHz, which equals the LCO frequency divided by 8, see Section IV-C) at the trimming temperature \( T_{\text{trim}} \) (measured on-chip, approximately RT but not strictly controlled), the trimming supply voltage \( V_{\text{trim}} \) and at time \( t_{\text{trim}} \). The 18-bit value of \( M_{\text{trim}} \) provides accuracy up to \( f_{\text{ext}}/(N_{\text{ext}} \cdot f_{\text{st}}) \approx 4 \text{ ppm} \) of \( f_{\text{st}} \). The implemented (non-ideal) LC-based \( f_{\text{st}} \) has a well-defined temperature dependency which is about \(-68.2 \text{ ppm/K} \) (see section V-B). To use the frequency sampler at a temperature other than \( T_{\text{trim}} \), the value of \( M_{\text{trim}} \) is compensated across temperature as

\[
M_{\text{st}}(T) = M_{\text{trim}} \cdot \left( 1 + \sum_{n=1}^{5} p_n \cdot \Delta T^n \right)
\]

In (4), \( \Delta T = T - T_{\text{trim}} \) and \( p_n \) are the coefficients of a 5\textsuperscript{th} order polynomial of the LCO frequency across temperature, determined by batch calibration.

B. CCO (re)calibration

For the CCO (re)calibration, the counter starts/stops by the scaled version of the LCO frequency \( f_{\text{st}}/N_{\text{st}} \) and counts the periods of \( f_{\text{ref}} \), yielding a counter value

\[
M_{\text{counter}}(P,V,T,L, \ldots) = \left[ f_{\text{st}}(P,V,T,L, \ldots) \cdot N_{\text{st}} \right] / f_{\text{ref}}(P,V,T,L, \ldots)
\]

During production trim, a target value for \( M_{\text{counter}} \) is set such that \( f_{\text{ref}} \) is equal to \( f_{\text{st}} \). For our demonstrator, \( f_{\text{st}} = 70 \text{ MHz} \). The target counter value \( M_{\text{tar}} \) is stored in the LUT as

\[
M_{\text{tar}}(T) = \left[ f_{\text{st}}(T) / N_{\text{st}}(T) \right]
\]

The division factor for \( f_{\text{st}} \) is set to \( N_{\text{st}} = 2^{17} = 1 \), which yields a resolution of 25 ppm in \( M_{\text{tar}} \) and hence also in \( f_{\text{ref}} \). The accuracy of \( M_{\text{tar}} \) across temperature depends on the temperature sensitivity of \( M_{\text{tar}}(T) \) (about \(-68.2 \text{ ppm/K} \), see section V-B) in combination with its temperature resolution. To have similar accuracy and resolution of \( M_{\text{tar}} \), the temperature resolution of the LUT is set to 0.5 °C, resulting in 470 LUT-values of \( M_{\text{tar}}(T) \) for the measured temperature range (see section V-B).

During (re)calibration, the digital core (see Figure 5) tunes \( f_{\text{ref}} \) to get \( M_{\text{counter}} = M_{\text{tar}}(T) \). The digital core consists of a subtractor and an error accumulator with gain \( a_{\text{ki}} \). The subtractor outputs \( M_{\text{tar}}(T) - M_{\text{counter}} \) which (binary) value is subsequently scaled by a factor \( a_{\text{ki}} \) and consequently accumulated. The accumulated error is then utilized as frequency control word for the CCO. The loop gain of the system, simplifying to a time-continuous loop without delay, is:

\[
f_{\text{ref}} / M_{\text{tar}} = a_{\text{ki}} \cdot a_{\text{CCO}} \cdot f_{\text{ref}} / N_{\text{st}} / \tau_{\text{lpf}} \cdot s^2 + \tau_{\text{lpf}} \cdot s + a_{\text{ki}} \cdot a_{\text{CCO}} \cdot a_{\text{ki}} \cdot N_{\text{st}}
\]
where \( \tau_{lpf} \approx 180 \mu s \) is the time-constant of the low-pass filter, \( \alpha_{CCO} \approx 300 \text{ Hz/code} \) is the combined gain of the digital-to-analog converter (DAC) and CCO, and \( \alpha_{is} = N_{is}/f_{st} \) is the gain of the frequency sampler. The dominant time constant of the loop is \( \tau \approx 1/(\alpha_{ki} \cdot \alpha_{CCO}) \approx 1.66 \text{ ms} \) for \( \alpha_{ki} = 2 \). The frequency sampler needs \( N_{is}/f_{st} \approx 525 \mu s \) for one conversion. To reduce an error of \( \epsilon_{freq} = 100 \text{ ppm in } f_{ref} \) to the \( M_{ref} \) resolution of 25 ppm takes the loop about \( \ln (100 \text{ ppm}/25 \text{ ppm}) \cdot \tau \approx 2.3 \text{ ms} \). This results in 5 complete frequency sampler conversions, yielding a total settling time of about 2.6 ms.

C. Colpitts oscillator

A modified version of the Colpitts oscillator from our work presented in [19] implements the LCO that generates the well-behaved frequency \( f_{st}(T) \) over PVTL for the frequency sampler. The oscillation frequency of the implemented Colpitts oscillator (see Figure 6) is [19]

\[
f_{LC} \approx \frac{1}{2\pi \sqrt{L \cdot C_{S}}} \times \sqrt{1 + \frac{1}{Q_{L}} \left( \frac{1}{Q_{C_L}} + \frac{1}{Q_{C_C}} \right)}
\]

where \( L \) is the tank inductance, \( C_{S} \) is the series capacitance of the tank capacitors \( C_{A} \) and \( C_{C} \), \( Q_{L} \), \( Q_{C_L} \) and \( Q_{C_C} \) are the quality factors of \( L \), \( C_{A} \) and \( C_{C} \), respectively.

A Colpitts oscillator achieves higher frequency stability across PVTL, in the low GHz range, than a cross-coupled LC-oscillator [19], [23]. For ideal reactances (no loss) both LCOs have the same temperature sensitivity of \( \partial f/\partial T \approx -T_{CL}/2 - T_{TC}/2 \) [19], where \( T_{CL} \) and \( T_{TC} \) are the TC of the inductance and capacitance, respectively. For finite quality factors, the temperature sensitivity of a cross-coupled- and a Colpitts LCO depend differently on the temperature-sensitive series resistance \( R_{s} \) of the inductor [19], [23]. For cross-coupled LCOs it can be derived that \( \partial f/\partial T \approx -T_{CR}/Q_{C}^2 \), where \( T_{CR} \) is the TC of \( R_{s} \). For a Colpitts oscillator, it can be derived that \( \partial f_{LC}/\partial T \approx 2T_{CR}/(Q_{C}Q_{C}) \) where \( Q_{C} \) is the quality factor of the tank capacitance \( C_{S} \). At oscillation frequencies where \( Q_{C} < Q_{C}^2 \) (the low GHz range) the Colpitts oscillator hence is intrinsically more stable over temperature than a cross-coupled oscillator.

For the current paper, we reused the design from [19] with a number of modifications. Firstly, to ensure that the impact of the Groszkowski effect [19], [34] is sufficiently small, we included amplitude control that regulates the oscillation amplitude to about 175 mV; this amplitude control is now fully integrated. Secondly, compared to the work in [19], the LCO frequency is slightly increased to 2 GHz and a different inductor layout is used, now having \( Q_{L} \approx 18 \) at RT and showing an estimated \( T_{L} \) of about 120 ppm/°C. Similar to the work in [19], a complementary-to-absolute-temperature bias voltage (CBV) is used for the LCO to compensate for the PVT dependency of the drain-bulk junction capacitance.

To assure proper operation of the control logic of the frequency sampler, a high speed prescaler divides the LCO frequency \( f_{LC} \) by 8 to generate \( f_{st} \).

D. Current-controlled oscillator

Figure 7 shows the block schematic of the programmable CCO that consists of an inverter-based 3-stage ring oscillator, a limiting amplifier and a low-pass filter, and the current DAC with a current reference circuit. The 9-bit current DAC is implemented by switchable current mirrors, where the 4 MSBs are implemented by thermometer code, and the 5 LSBs are implemented binary weighted. The 9-bit current DAC delivers the supply current for the CCO core, which is mirrored and low-pass filtered via \( R_{LPF} \) and \( C_{LPF} (f_{3.6dB} \approx 860 \text{ Hz}) \) by the PMOS mirror; its smallest current step, before applying \( \Sigma \Delta \) modulation, is about 125 nA which corresponds to a frequency step of 77 kHz. The \( \Sigma \Delta \) modulator (see Figure 5) adds 8 fractional bits of resolution to the DAC by duty-cycling an LSB-step across 255 cycles. This effectively achieves...
17 bits of resolution. The frequency resolution of the CCO is thereby increased to 300 Hz, or 4 ppm of \( f_{\text{ref}} \). The \( \Sigma \Delta \) modulator operates at 70 MHz/32 ≈ 2.2 MHz, and has a minimal repetition frequency of 8.5 kHz, which is about an order of magnitude higher than the corner frequency of the succeeding low-pass filter. The operating frequency of the \( \Sigma \Delta \) modulator is therefore chosen sufficiently low to achieve relatively low \( \Sigma \Delta \) power consumption while still achieving the target resolution and is high enough to not affect the calibration speed.

### E. Temperature sensing

The temperature sensor continuously monitors the die temperature for the digital calibration engine to select the proper \( M_{\text{trim}} \) value from the LUT and to start (re)calibration when deemed necessary. Absolute accuracy of the temperature sensor is essential for selecting the correct LUT data (0.5°C steps) that is used to compensate the LCO frequency drift across temperature (about −68.2 ppm/°C, see section V). During trimming, the LUT is generated from \( M_{\text{trim}} \) (see Section IV-A), the simultaneously sensed die temperature \( T_{\text{trim}} \) and batch information. Static inaccuracies of the temperature sensor are hence accounted for in the LUT data and have negligible impact on the system performance. The quantization resolution of the temperature sensor must be better than 0.4°C for triggering the (re)calibration before \( f_{\text{ref}} \) deviates by more than \( \epsilon_{\text{ref}} \) (see section III-A) due to the CCO temperature coefficient, which is about 242 ppm/°C (see Section V-B). It is desired that the rms temperature resolution is significantly higher than the quantization resolution of 0.4°C to avoid triggering recalibrations due to noise, which would lead to increased power consumption.

Figure 8 shows the block diagram of the temperature sensor, consisting of a temperature sensing front-end that generates two temperature-dependent currents, \( I_{\text{PTAT}} \) and \( I_{\text{CTAT}} \), and a dual-slope ADC quantizing the temperature-dependent ratio \( X(T) = I_{\text{PTAT}} / I_{\text{CTAT}} \). The counter of the dual-slope converter is driven by an external clock \( f_{\text{clk}} \) of 10 MHz for measurement flexibility. However, this clock could be derived from the CCO clock as well. The dual-slope architecture is insensitive to static changes in the clock frequency. The currents \( I_{\text{PTAT}} \) and \( I_{\text{CTAT}} \) are generated in the BTBT-based sensing front-end from the base-emitter voltage difference \( \Delta V_{\text{BE}} = (kT/q) \ln(n) \) and the base-emitter voltage \( V_{\text{BE,1}} \), respectively. The currents \( I_{\text{PTAT}} \) and \( I_{\text{CTAT}} \) are then obtained by \( V \)-to-1 conversion by resistors \( R_1 \) and \( R_2 \) to generate \( I_{\text{PTAT}} = \Delta V_{\text{BE}} / R_1 \) and \( I_{\text{CTAT}} = V_{\text{BE}} / R_2 \), respectively, similar to [35]. The ratio \( X(T) \), which is digitized by the dual-slope converter, can be expressed by

\[
X(T) = \frac{I_{\text{PTAT}}}{I_{\text{CTAT}}} = \frac{R_1}{R_2} \cdot \frac{V_{\text{BE,1}}}{\Delta V_{\text{BE}}} \tag{8}
\]

Note that \( X(T) \) is largely independent of the temperature coefficient of resistors \( R_1 \) and \( R_2 \). Dynamic element matching (DEM) and chopping of the amplifiers and current mirrors are implemented to reduce the effect of process spread and flicker-noise to obtain sufficient sensor accuracy. The digitized ratio \( \mu_{\text{out}} = r \cdot [X(T)] \), where \( r \) is a gain-factor of the dual-slope converter, is used off-chip to derive the digital output \( D_{\text{out}} \) (proportional to the temperature \( T \)) similar as [36]. The final value of \( D_{\text{out}} \) is the average of six sub-conversions \( D_{\text{out,1}}, \ldots, D_{\text{out,6}} \), which use different settings for the DEM and chopping switches. A sample-specific 1T-trim and a 3rd-order polynomial based on batch calibration (see section V-A) are used to correct non-linearities of the temperature sensor.

### F. Supply regulation

The supply regulation is implemented by two low-dropout regulators (LDOs) that generate the 2.5 V and 1.5 V supply, respectively. The LDOs have no tuning circuitry, but share a reference voltage from a bandgap circuit that is tuned to its nominal value, of about 1 V, at \( T_{\text{trim}} \). The TC of the bandgap circuit is corrected by batch calibration. The (measured) supply voltages have uncorrelated spread in their output voltage of ±50 mV (3σ) and ±40 mV (3σ) for the 2.5 V and 1.5 V supply, respectively. The 1.5 V LDO drives the temperature sensor, the CCO with its current DAC and \( \Sigma \Delta \)-modulator, the digital circuitry and some parts of the LCO. The duty-cycled Colpitts oscillator is supplied via the 2.5 V LDO.

### V. Measurement Results

A prototype in a 0.13 \( \mu \)m HV CMOS SOI process is fabricated and packaged in a plastic package (transfer molding process). Figure 10 shows a die micrograph of the prototype. The prototype is not optimized for area and occupies a combined active area of 0.69 mm². In this paper, the temperature sensor and LCO (including LDO) were characterized in separate measurements to obtain batch information optimized for temperature stability and measurement speed, respectively. This characterization could be done at once in a production scenario. The final 1T-trim of all sub-blocks, complementary to the batch calibration, is carried out simultaneously.
A. Characterization of the temperature sensing circuit

The temperature sensing circuit of 6 samples is characterized across a temperature range from −40 °C to 130 °C (measurement setup limited) to obtain information for batch calibration. The measurements were executed in a climate chamber, and the samples were in good thermal contact with a PT100 (resistance thermometer) for reference measurement. Figure 11 shows the measured temperature error after applying a single-point trim and using the obtained batch calibration, for the 6 samples. The measured span of the temperature error across the 170 °C temperature range is about 0.3 °C. The conversion-time of the temperature sensor is about 1.2 ms and it consumes about 46.2 µW from the 3.3 V supply at RT. The rms-resolution is 25 mK, corresponding to a resolution FoM [37] of 34.8 pJ · K².

B. Uncalibrated/untrimmed CCO and LCO frequency

The uncalibrated/untrimmed frequency of the CCO and LCO is characterized across a temperature range from −65 to 170 °C provided by a thermal-streamer (TA-5000). The corresponding junction temperature measured by the temperature sensor is −63 to 165 °C; the difference is due to RT airflow and thermal leakage. In all measurements this
junction-temperature is used as temperature reference for the reported performance.

Figure 9(a) shows the measured CCO frequency of 18 samples. For demonstration purposes, all samples use the same, static, tuning value, showing the large spread in both frequency and temperature behaviour. Figure 9(b) shows a measure for the LCO frequency, that is measured via the digital frequency sampler output as \( f_{sd} = f_{LC}/8 \approx M_{counter} \cdot f_{ext}/N_{ext} \). The measured frequency deviation w.r.t. the sample-specific frequencies at \( T_{\text{lim}} \) are shown in Figure 9(c) and (d) for the CCO and LCO, respectively. The CCO frequency drifts with roughly 242 ppm/°C (box-method for the measured temperature range) and shows significant TC-spread over the measured samples. The LCO frequency deviation stays within ±0.8% across the temperature range, but more importantly, the TC of −68.2 ppm/°C (box-method) is consistent over the 18 measured samples, as is shown in Figure 9(d). Based on the first 11 samples measured, a 5th-order polynomial \( p(T) \), which models the LCO temperature behavior, is obtained for batch calibration.

The maximum frequency error between \( p(T) \) and the measured samples is about ±75 ppm, which includes errors from the temperature sensor. A larger error is expected for wafer-to-wafer spread [19]. If a 3rd-order polynomial \( p(T) \) would be used instead of the 5th-order polynomial, the maximum frequency error between the polynomial and the measured samples would increase to about ±120 ppm.

C. Calibrated CCO frequency

As described in section IV-B, the frequency sampler is trimmed at RT with an external clock of 1 MHz and the result together with the batch calibrated polynomial \( p(T) \) is used to generate the LUT data. After this, the samples are measured from −65 °C to 170 °C (ambient). At each measurement point the CCO is recalibrated as described in section IV-B. Figure 12 shows the calibrated CCO frequency (left y-axis) and frequency inaccuracy w.r.t. 70 MHz (right y-axis) of 18 samples as a function of temperature.

The maximum frequency error between \( p(T) \) and the measured samples is about ±75 ppm, which includes errors from the temperature sensor. A larger error is expected for wafer-to-wafer spread [19]. If a 3rd-order polynomial \( p(T) \) would be used instead of the 5th-order polynomial, the maximum frequency error between the polynomial and the measured samples would increase to about ±120 ppm.

D. Power consumption

When the system is not recalibrating, and hence in low-power mode, only the 1.5 V voltage regulator, the temperature sensor, the CCO and the associated ΣΔ modulator are powered. At room temperature, these circuits draw a total of ≈64 µA from the 3.3 V supply. Figure 14 shows the power breakdown in low-power mode. While the system is recalibrating also the 2.5 V voltage regulator and frequency sampler (including the Colpitts oscillator) are powered which draw an additional 4.75 mA from the 3.3 V supply.

E. Time domain measurement

Fast (re)calibration of the CCO is necessary to minimize the on-time of the LCO and hence lower the duty-cycle and the overall power consumption. The start-up time of the LCO, including start-up of the accompanying LDO and settling of the amplitude control loop, is less than 30 µs. Figure 15 shows transient measurements of the CCO frequency during an operational recalibration cycle as described in section IV-B. For demonstration purposes we measured the recalibration speed as function of the gain setting \( k_i \). For this measurement, the CCO frequency \( f_{ref} \) was manually set to about 69 MHz instead of the desired 70 MHz (~15 000 ppm error) before recalibration was started. The curves in Figure 15 correspond...
to the measured $f_{\text{ref}}$ over recalibration time for various $a_{ki}$. These correspond to different time constants as described by (7). To achieve both fast settling and low residual error when stopping the (re)calibration loop, the gain can be geared down during recalibration (the curve denoted by $a_{ki}=2.00 \rightarrow 0.25$).

The measurement resolution of the trimming and calibration loop depends both on the counter- and LUT values ($M_{\text{counter}}$ and $M_{\text{LUT}}(T)$) and on the Allan deviation of the clocks. Figure 16 shows the measured Allan deviation of the LCO and CCO without a strictly controlled environmental temperature and without intermediate calibrations. The measured Allan deviation floor, which limits the accuracy of trimming and calibration, is approximately 0.7 ppm for the LCO and 25 ppm for the CCO. Our demonstrator is optimized for high PVTL stability, and not for (noise limited) short-term output frequency stability of the CCO. There is, however, no fundamental reason why this could not be improved further using e.g. the approach in [11]–[13]. The rms period jitter of the free-running CCO in an uncontrolled environment is measured to be 14.5 ps, as shown in Figure 17. The measured phase noise of the free-running CCO is shown in Figure 18.

**F. Frequency deviation over lifetime**

To evaluate the sensitivity of the system over lifetime effects, 5 samples have undergone a 6-day storage bake at 175 °C, which corresponds to 150 years at 25 °C [38]. A storage bake is a reasonable representation of the aging scenario if the frequency sampler has a very small duty cycle.

Figure 19 shows the measured output frequency deviation over temperature before and after the bake. The mean open-loop CCO frequency with the same manual trim code as in section V-B shows a large shift of approximately −2500 ppm (as expected from section II), while the mean CCO frequency after calibration with the frequency sampler has shifted only about +52 ppm, which is mainly attributed to a similar shift in the LCO frequency. This is in accordance with results in [1], [18], [19]. This indicates that the presented frequency reference is much less sensitive to aging than a plain (free-running) CCO.

©2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Table I

<table>
<thead>
<tr>
<th>reference principle</th>
<th>This work</th>
<th>[18]</th>
<th>[20]</th>
<th>[7]</th>
<th>[11]</th>
<th>[12]</th>
<th>[13]</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency [MHz]</td>
<td>70</td>
<td>100</td>
<td>25</td>
<td>16</td>
<td>28</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>temperature range [°C]</td>
<td>-63 to 165</td>
<td>-40 to 140</td>
<td>-10 to 80</td>
<td>-55 to 125</td>
<td>-40 to 85</td>
<td>-45 to 85</td>
<td>-45 to 85</td>
</tr>
<tr>
<td>TC [ppm/°C] (# temperature trimming points)</td>
<td>0.7 (Batch+1)</td>
<td>0.5  (Batch+2)</td>
<td>2.9  (Batch+1)</td>
<td>11.1  (Batch+1)</td>
<td>2.6  (Batch+2)</td>
<td>5.2  (Batch+1)</td>
<td>1.3  (Batch+2)</td>
</tr>
<tr>
<td>supply sensitivity [ppm/V]</td>
<td>92×</td>
<td>2.6×</td>
<td>20×</td>
<td>NA</td>
<td>2000</td>
<td>2000</td>
<td>1200</td>
</tr>
<tr>
<td>number of samples</td>
<td>18</td>
<td>28</td>
<td>10</td>
<td>24</td>
<td>12</td>
<td>18</td>
<td>20</td>
</tr>
<tr>
<td>power [mW]</td>
<td>0.21b</td>
<td>14.85c</td>
<td>59.4</td>
<td>2.1</td>
<td>0.14</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>Allan deviation floor [ppm]</td>
<td>25 (@ 1 s)</td>
<td>NA</td>
<td>26 (@ 300 s)</td>
<td>NA</td>
<td>2 (@ 40 s)</td>
<td>0.35 (@ 100 s)</td>
<td>0.32 (@ 11 s)</td>
</tr>
<tr>
<td>area [mm²]</td>
<td>0.69</td>
<td>0.2c</td>
<td>NA</td>
<td>0.5</td>
<td>0.06</td>
<td>0.14</td>
<td>0.3</td>
</tr>
<tr>
<td>technology</td>
<td>130 nm</td>
<td>130 nm</td>
<td>250 nm</td>
<td>160 nm</td>
<td>65 nm</td>
<td>180 nm</td>
<td>180 nm</td>
</tr>
</tbody>
</table>

NA = Not Available  a Includes an on-chip LDO.  b Includes only always on-power, which assumes a small duty cycle (<0.03%) and hence negligible power overhead of the (re)calibration system.  c From private communication as not available from [18] itself. Power is calculated for a supply voltage of 2.7 V.  d To the best of the authors knowledge, there is no specific benefit of using an HV SOI process for this system.

G. Benchmarking

Table I summarizes the measured performance of the presented design and compares it to other fully integrated frequency references. The achieved frequency accuracy over temperature of this work is more than 4 times better compared to the LC-based work of [20] and 7 times better than the RC-based work of [12], which also only use a 1T-trim. Moreover, we report the largest temperature range and demonstrate a high stability over lifetime which makes the system suitable for the most stringent automotive temperature requirements. The 210 µW power consumption is on par with the RC-based work in [11]–[13] despite our higher output frequency. Simultaneously, this work achieves better frequency accuracy and integrates LDOs and a temperature sensor. Due to the integrated supply regulation our work achieves more than one order of magnitude better supply sensitivity compared to the RC-based references [11]–[13]. The RC-based references in [11]–[13], that also tune a ring-oscillator to generate the output frequency, have a lower Allan deviation floor due to their continuously operating PLL.

VI. CONCLUSIONS

A fully-integrated frequency reference system in a 0.13 µm HV CMOS SOI process has been presented. The system effectively combines the high frequency accuracy over PVT of a best-in-class untuned Colpitts LCO with a low-power but inherently highly PVT sensitive RC-based CCO. The low-power CCO is periodically, over temperature, recalibrated to 70 MHz with the highly accurate duty-cycled LCO. Our demonstrator achieves frequency accuracies better than ±93 ppm over a temperature range from −63 to 165 °C (resulting in 0.7 ppm/°C) across 18 samples while requiring only a single room-temperature trim. The mean frequency shift due to a storage bake is only +52 ppm. The presented frequency reference has the best reported combined temperature coefficient and power efficiency (3 µW/MHz) compared to state-of-the-art 1T-trimmed fully-integrated frequency references.

VII. ACKNOWLEDGMENTS

We thank F. de Vries and N. Wils, both of NXP Semiconductors, for technical support and measurement support, respectively. We also thank A. Rop from the University of Twente for lab support.

REFERENCES

Anne-Johan Annema (M’00) received the M.Sc. degree in electrical engineering and the Ph.D. degree from the University of Twente, Enschede, The Netherlands, in 1990 and 1994, respectively. In 1995, he joined the Semiconductor Device Architecture Department, Philips Research, Eindhoven, The Netherlands, where he was involved in a number of physics-electronics-related ranging from low-power, low-voltage circuits, fundamental limits on analog circuits related to process technologies, and high voltage in baseline CMOS to feasibility research of future CMOS processes for analog circuits. Since 2000, he has been with the IC Design Group, Department of Electrical Engineering, University of Twente, where he is Associate Professor. He is also a part-time Consultant in industry and a Co-Founder of ChipDesignWorks. His current research interests include physics, analog, and mixed-signal electronics, RF power and deep submicrometer technologies, and their joint feasibility aspects.

Yanyu Jin received the B.Eng. degree in electronic and information engineering from Zhejiang University, Hangzhou, China, in 2003, the S.M. degree in high performance computation in engineered systems from National University of Singapore in 2004, and the M.S. degree in electrical engineering from Delft University of Technology (TU Delft), the Netherlands, in 2006, where he continued as a research assistant, working on mm-wave transceiver building blocks for integrated wireless applications. Mr. Jin joined NXP Semiconductors, the Netherlands, in 2011, and has since worked on RF MIC, analog and mixed-signal IP development. His current research interests include crystal-less high-accuracy frequency generation in CMOS.

Jos Verlinden received his M.Sc. degree in electrical engineering from the Eindhoven University of Technology, in 1997, The Netherlands. In 1997, he joined the central organization of research and development at Philips Semiconductors, Eindhoven, the Netherlands, where he developed analog mixed signal IPs for different applications. From 2010 to 2015, he has been working on the design of integrated RF transceiver solutions. Since 2015 he is cluster architect in the central AMS IP group from NXP Semiconductors where he is responsible for the development of, amongst others, V/I references and clocking solutions for different applications.

Bram Nauta (M’91–SM’03–F’08) was born in Hengelo, The Netherlands, in 1964. He received the M.Sc. degree (cum laude) in electrical engineering and the Ph.D. degree in analog CMOS filters for very high frequencies from the University of Twente, Enschede, The Netherlands, in 1987 and 1991, respectively. In 1991, he joined the Mixed Signal Circuits and Systems Department of Philips Research, Eindhoven, The Netherlands. In 1998, he returned to the University of Twente, where he is currently a Distinguished Professor, heading the IC Design Group. Since 2016, he has been serving as the Chair of the EE Department, University of Twente. His current research interests are highspeed analog CMOS circuits, software-defined radio, cognitive radio, and beamforming. He is also a member of the Royal Netherlands Academy of Arts and Sciences (KNAW). He was a co-recipient of the ISSCC 2002 and 2009 “Van Vessem Outstanding Paper Award.” In 2014, he received the Simon Stevin Meester Award ($500,000), the largest Dutch national prize for achievements in technical sciences. In 2013, he was the Program Chair of the International Solid State Circuits Conference (ISSCC). He served as the President of the IEEE Solid-State Circuits Society (2018–2019 term). He has served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II from 1997 to 1999 and JSSC from 2001 to 2006. He has served as the Editor-in-Chief of the IEEE Journal of Solid-State Circuits (JSSC) from 2007 to 2010. He was on the Technical Program Committee of the Symposium on VLSI circuits from 2009 to 2013 and on the Steering Committee and the Programme Committee of the European Solid State Circuit Conference (ESSCIRC). He has served as the Distinguished Lecturer of the IEEE.