All-digital SoC Thermal Sensor using On-chip High Order Temperature Curvature Correction

Mehdi Saligane, Mahmood Khayatzadeh, Yiqun Zhang, Seokhyeon Jeong,

David Blaauw, Dennis Sylvester

University of Michigan, Ann Arbor, MI, USA

***Abstract* — Accurate, compact thermal sensors are desirable in many applications, including on-chip temperature monitoring for processors with dynamic throttling and reliability management. Modern thermal sensors are limited in either area, robustness, or accuracy. This work sidesteps strong linearity requirements for reference and PTAT elements in the sensor by performing a higher-order fitting of more relaxed PTAT and CTAT elements using an embedded calculation compute engine. A compact 24 × 10μm sensing element (40nm CMOS) achieves inaccuracy of <1°C across 30 chips with 2-point calibration and a resolution of 0.02°C.**

I. Introduction

Embedded thermal sensors aid in detecting hotspots, which are often difficult to anticipate during the design phase [1-3]. Increasing sensor count across the die improves thermal map granularity, but sensor size becomes limiting. Various approaches to design accurate, area-efficient sensors have been proposed [1-5, 7, 9-10]. The thermal monitor in [2] is compact with good accuracy over a significant temperature range but does not include readout circuitry. The sensor in [3] achieves high accuracy over a wide temperature range but relies on a BJT, which is not widely available in CMOS processes and requires a large analog-to-digital converter. Other conventional thermal sensors are often diode-based and while they have low process dependency and voltage sensitivity, they offer poor temperature accuracy (e.g., 3.1°C [4]). Also, sensors based on temperature-insensitive references [1, 5-6] can be difficult to calibrate (Fig. 1) and are prone to noise from core logic toggling.

II. Proposed Temperature Sensor

This work proposes a fully synthesizable, compact, and accurate temperature sensor for fine-grain temperature monitoring in high performance processors and other area-constrained, advanced process technology applications. Using simple process information and two-point temperature calibration, a high-order polynomial fit of the sensing elements is performed. The two polynomial functions are inverted and solved using a small co-located compute engine to produce a corrected, linearized temperature code. Using higher-order (quadratic and cubic) fitting allows the linearity requirement of the sensing elements to be greatly relaxed, which results in significant area and power improvements. This approach shifts the burden of linearity from the analog sensing elements (Fig. 1, conventional approach) to digital computation, which takes advantage of advanced process technologies where analog design is challenging and digital computation is energy efficient (Fig. 2).

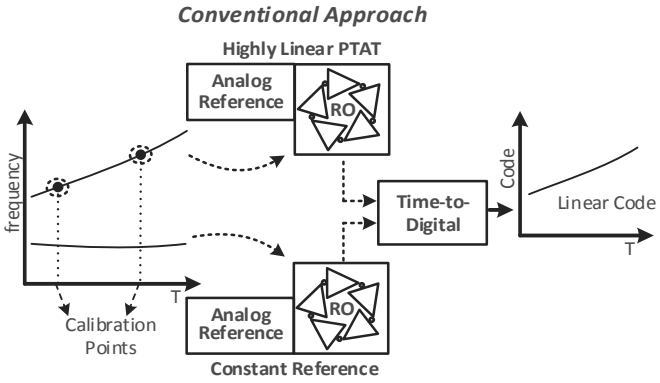

Fig. 1. Conceptual diagram of a conventional temperature sensor.

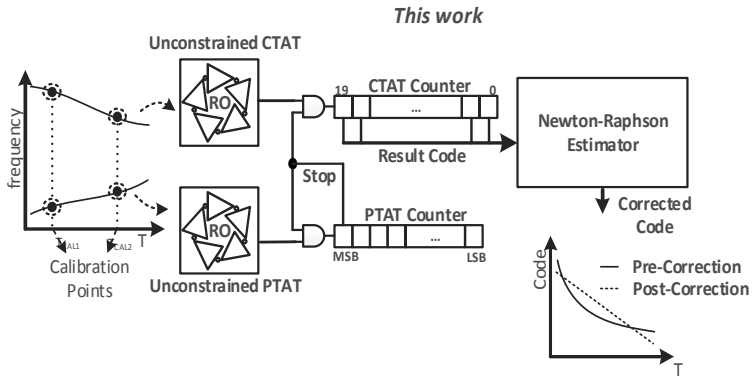

Fig. 2. Conceptual diagram of the proposed temperature sensor.

978-1-4799-8682-8/15/$31.00 ©2015 IEEE

The co-located compute engine can be shared among multiple sensors or replaced by the host processor. We present two implementations: 1) a stand-alone sensor with co-located compute engine and 2) an ARM R4 processor with 20 embedded sensors that use the main CPU for digital processing. The sensor achieves <1°C accuracy across -40°C to 100°C with a sensing element of only 240μm^2 in 40nm CMOS. The implemented sensor uses two ordinary ROs without compensation, reducing design complexity (Fig. 3). To obtain high resolution, the frequency of one RO is complementary to temperature (CTAT) and the second is proportional (PTAT), which is achieved by operating the ROs at nominal (1V) and low (0.5V) supply voltages, respectively.

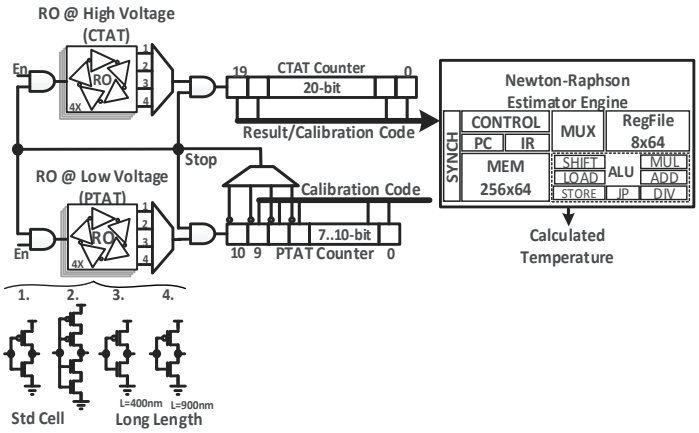

Fig. 3. Circuit diagram of the standalone temperature sensor.

A configurable 8-11b PTAT counter generates a temperature-sensitive reference time. The MSB of this counter stops both ROs and captures a result code in a 20-bit counter. This code is proportional to the ratio of CTAT to PTAT frequencies (Fig. 2) and is inversely proportional to temperature and highly non-linear. However, each of the two frequencies (F_{CTAT} and F_{PTAT}) are fit during a one-time 2-point calibration process with a polynomial function and the resulting coefficients (a_k and b_k) are stored. Equation (1), expressed in terms of these coefficients and the result code (Equation (2)), is solved with a Newton-Raphson estimator using the compute engine to obtain a corrected code that is highly linear with temperature:

$$P(T) = \sum_{k=0}^{3} \left(a_k - \frac{b_k . Code_{Res}}{2^N - 1} \right) . T^k \qquad (1)$$

Where:

$$Code_{Res} = \left(2^N - 1\right) . \frac{F_{CTAT}}{F_{PTAT}} \qquad (2)$$

Solving:

$$P\left(T_{Corrected}\right) = 0 \qquad (3)$$

The compute engine contains a 64b fixed-point ALU and a 256×64 memory and operates at 47MHz at 0.5V (measured).

III. Fitting Methodology

As we will show, the use of simple ROs requires 3rd order polynomial (cubic) fitting in order to achieve 1°C accuracy. However, such a fit requires the two RO frequencies to be measured at four temperatures for each die, leading to excessive testing time. To maintain standard two-point calibration, only the 1st and 2nd coefficients (a_0, a_1, b_0, b_1) are fit based on frequency measurements while 3rd and 4th coefficients (a_2, a_3, b_2, b_3) are approximated using global process information as follows. One slow and one fast die, as indicated by one of the RO frequencies, are selected from among a large set of chips and characterized across the entire temperature range (Fig. 4).

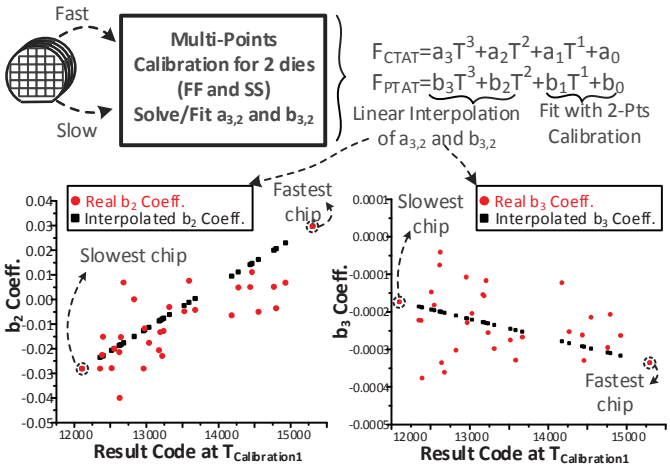

Fig. 4. Fitting methodology using linear interpolation of $a_{3,2}$ and $b_{3,2}$.

Fourth-order fitting is then performed (only for these two die) and the resulting 3rd and 4th coefficients recorded. During calibration of all other die, a linear interpolation based on the RO frequency of that particular die is then used to set its two higher-order coefficients. In this way the higher-order coefficients, which have a relatively small weight, are approximated while the lower-order coefficients are accurately fit using a standard two-point calibration process.

IV. Embedded Temperature Sensor

In the embedded implementation, the sensing elements are uniformly distributed across an ARM R4 processor. The small size of the sensor allows fine-grain temperature monitoring and minimum invasiveness.

Fig. 5 shows the implemented R4 core with 20 embedded sensors as well as the measured temperature gradient across the core during periods of different activity. The recorded temperature of an embedded sensor during a transition between idle and high activity codes (FFT and prime number calculations) is also shown. Although the temperature change is small in this low-power medium-size core, the measured temperature accurately tracks workload change and shows good spatial resolution.

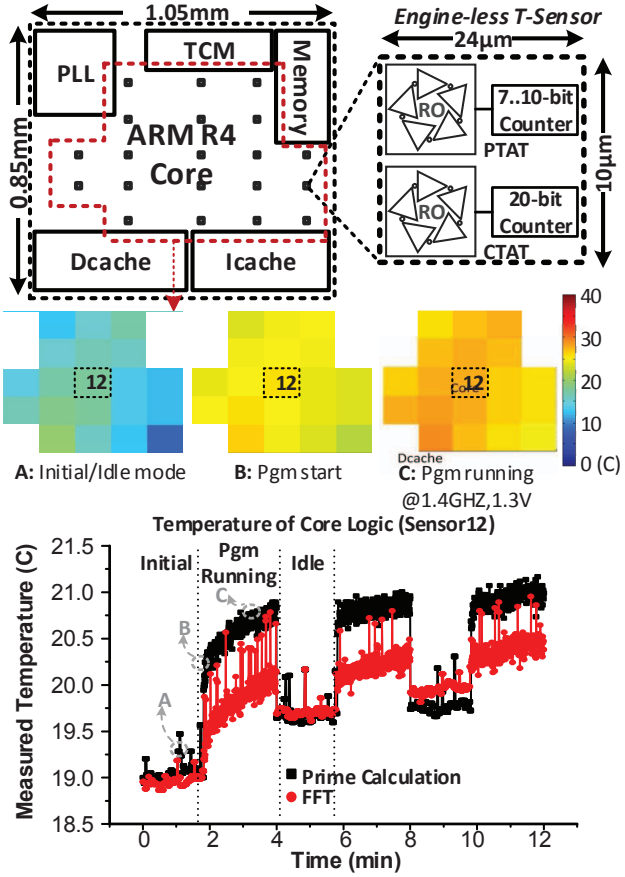

Fig. 5. Temperature map of ARM R4 Processor and temperature evolution for different workloads.

V. Measurement Results

The proposed sensor and calibration technique were implemented in 40nm CMOS and tested on 30 die. Fig. 6 shows a measured accuracy of -0.95°C / 0.97°C using a cubic fit with the proposed two-point calibration over a wide temperature range (-40 to 100°C). As expected, quadratic and linear fitting show somewhat decreased accuracy of -1.37°C / 1.29°C and -2.86°C / 1.91°C, respectively.

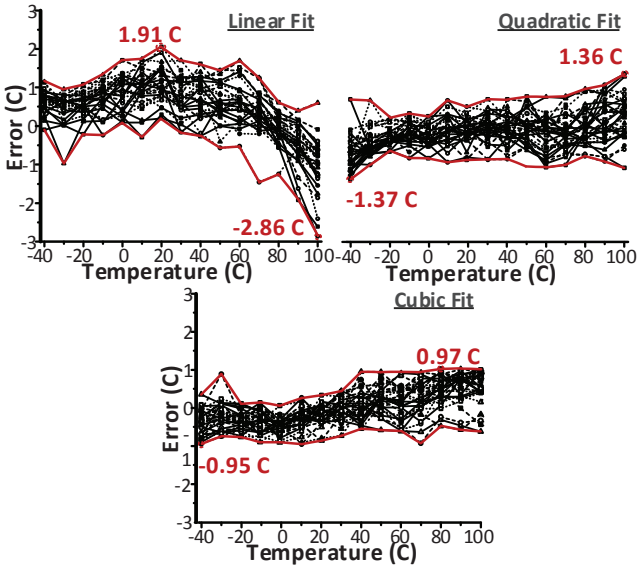

Fig. 6. Inaccuracy measurements across 30 chips using linear, quadratic, and cubic fits.

The standalone temperature sensor consumes 241µW at room temperature; this power is dominated by the compute engine. This can be circumvented by using a small number of cycles in the host processor to run the fitting algorithm, as done here on the ARM R4 implementation. In this case power consumption is 17µW per sensor. Additionally, the sensor's fast conversion time allows for the averaging of many samples at a small energy cost. As illustrated in Fig. 7, a potentially high standard deviation due to the noisy environment within a running processor can be greatly reduced with a small number of successive temperature measurements.

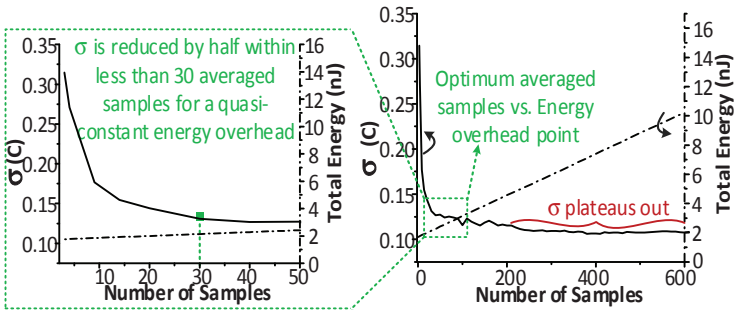

Fig. 7. Measured results showing how averaging can be used to improve measurement error (right), and a zoom-in on the optimum averaging vs. energy overhead point (left).

TABLE I

COMPARISON TABLE OF EMBEDDED TEMPERATURE SENSORS

Parameters	This work Embedded	[2]	[3]		[4]	[5]
Technology	40nm	65nm	32nm	22nm	65nm	0.13µm
Type	CMOS	CMOS	BJT		CMOS	CMOS
Area (μm^2)	240	279*	20000	6100	1225**	120000
Supply (V)	0.5~1	0.6~1	1.4	1.35	1~1.2	1.2
T-Range (°C)	-40~100	0~100	20~100		40~90	0~100
Resolution	0.02°C$_{rms}$	NA	0.19°C$_{rms}$	0.25°C$_{rms}$	1°C	0.66°C
Conversion Time (ms)	0.02	NA	0.5	0.7	1	0.2
Calibration	2-point	2-point	2-point		NA	1-point
Inaccuracy (°C)	0.97/-0.95[1a]	1.5/-1.6	1.2/-1.2	1.5/-1.5	3.1[1c]	2.3/-1.8
Power (µW)	17[1b]	0.92	3780	1390	NA	1200
Energy/ Conversion	0.34nJ[1b]	NA	19µJ	0.97µJ	NA	0.24µJ

1a. Using cubic fitting 1b. One SoC sensor 1c. Max. error

1d. Include the averaging of 200 samples

1e. FOM[nJ*K^2]=energy/conversion x Resolution^2

1f. Include energy/power of compute engine

*. Not including analog to digital circuitry

**. Thermal diode area not included

***. Power/Energy for generating external references is not included

Tables I and II summarize the specifications and performance of both the embedded and standalone sensors and compare them with recent work. The sensor shows fast conversion time, low energy per conversion, and wide temperature range along with good accuracy, small area, and high design portability. Die photos are given in Fig. 8.

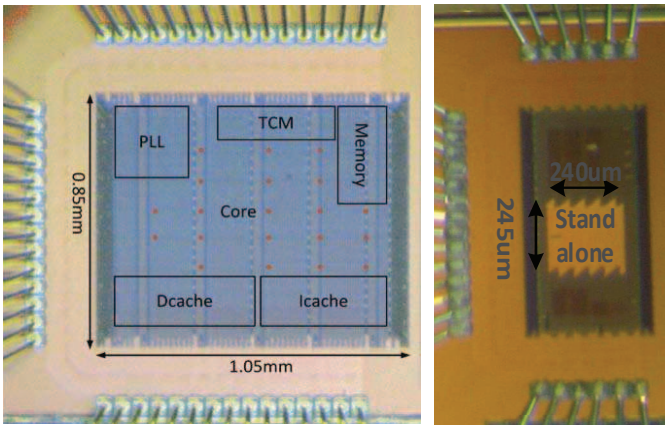

Fig. 8. Die photos of the distributed temperature sensor across an ARM R4 processor (left), and standalone temperature sensors (right).

TABLE II

COMPARISON TABLE OF STANDALONE TEMPERATURE SENSORS

Parameters	This work Standalone	[6]	[7]	[8]	[9]	[10]
Technology	40nm	700nm	0.16µm	0.18µm	65nm	0.18µm
Type	CMOS	BJT	CMOS	CMOS	CMOS	CMOS
Area (mm^2)	0.058	1.5	0.085	0.18	0.008	0.09
Supply (V)	0.5~1	2.9~5.5	0.85~1.2	1.2	1	1.2
T-Range (°C)	-40~100	-45~130	-40~125	0~100	0~110	0~100
Resolution	0.02°C$_{rms}$	0.003	0.063°C$_{rms}$	0.25°C	0.94°C$_{rms}$	0.3°C$_{rms}$
Conversion Time (ms)	0.02	2.2	6	0.0125	0.00213	30
Calibration	2-point	1-point	1-point	1-point	1-point	2-point
Inaccuracy (°C)	0.97/-0.95[1a]	+/-1.5	+/-0.4	0.5/-0.5	1.5/-1.4[1c]	1.5/-1.4
Power (µW)	241	159.5	0.6***	24***	500	71
Relative Inaccuracy	0.69	0.2	0.48	1	2.7	2.9
Energy/ Conversion	4nJ[1d/1f]	350nJ	3.6nJ***	0.3nJ***	1.1nJ	2.2nJ
FOM[1e]	0.0016	0.0032	0.014***	0.019***	0.94	0.19

ACKNOWLEDGEMENT

The authors thank TSMC University Shuttle Program for chip fabrication and acknowledge NSF for their support.

REFERENCES

[1] Poki Chen et al, "A Time-Domain SAR Smart Temperature Sensor With Curvature Compensation and a 3 Inaccuracy of •0.4°C ~ +0.6°C Over a 0°C to 90°C Range," JSSCC, March 2010.

[2] Teng Yang et al "16.4 0.6-to-1.0V 279µm2, 0.92µW temperature sensor with less than +3.2/-3.4°C error for on-chip dense thermal monitoring," ISSCC, Feb. 2014.

[3] J. Shor, et al., "Ratiometric BJT-Based Thermal Sensor in 32nm and 22nm Technologies," ISSCC, Feb. 2012.

[4] Saneyoshi, Eisuke, et al, "A 1.1V 35µm × 35µm thermal sensor with supply voltage sensitivity of 2°C/10%-supply for thermal management on the SX-9 supercomputer," VLSI, June 2008.

[5] K. Woo et al., "Dual-DLL-based CMOS all-digital temperature sensor for microprocessor thermal monitoring," ISSCC, Feb. 2009.

[6] Heidary, A et al, "12.8 A BJT-based CMOS temperature sensor with a 3.6pJ·K2-resolution FoM," ISSCC, Feb. 2014.

[7] Souri, K et al, "12.7 A 0.85V 600nW all-CMOS temperature sensor with an inaccuracy of ±0.4°C (3) from •40 to 125°C," ISSCC, Feb. 2014.

[8] Chun-Kuan Wu et al, "A 80kS/s 36µW resistor-based temperature sensor using BGR-free SAR ADC with a unevenly-weighted resistor string in 0.18µm CMOS," VLSI, June 2011.

[9] Hwang, S. et al, "A 0.008 mm2 500 µW 469 kS/s frequency-to-digital converter based CMOS temperature sensor with process variation compensation". TCAS, 2013.

[10] Chun-Kuan Wu et al, "A 80kS/s 36µW resistor-based temperature sensor using BGR-free SAR ADC with a unevenly-weighted resistor string in 0.18µm CMOS," VLSI, June 2011.