An Area-Efficient Microprocessor-Based SoC With an Instruction-Cache Transformable to an Ambient Temperature Sensor and a Physically Unclonable Function

Jiangyi Li, *Student Member, IEEE*, Teng Yang, *Student Member, IEEE*, Minhao Yang, *Member, IEEE*, Peter R. Kinget, *Fellow, IEEE*, and Mingoo Seok, *Member, IEEE*

***Abstract*— This paper presents a novel microprocessor-based system-on-chip (SoC) with the capability to transform SRAM in the instruction cache to an ambient temperature sensor and a physically unclonable function (PUF). For the transformation, we extend the original microprocessor without interlocked pipeline stages instruction set architecture and update SRAM peripheral circuits and pipeline control. All the changes are made minimally to reduce hardware overhead and the impact on the original microarchitecture. Compared to the conventional approach implementing dedicated hardware for low duty-cycle sensing and PUF, the proposed transformation approach saves ∼9.8× silicon area while achieving the performance and robustness comparable to the state of the art in implementing those functions. The efficiency of the approach is verified with an SoC prototyped in a 65-nm CMOS.**

***Index Terms*— Area efficient design, Internet of Things (IoT), physically unclonable function (PUF), system on chip (SoC), temperature sensor.**

I. Introduction

HEADING toward the era of Internet of Things (IoT), it is critical for integrated-circuit research and development to deliver compact, low cost, and dependable edge devices with various capabilities, e.g., sensing, computing, communication, and security [1]. This challenge has motivated to integrate an increasing number of components and function blocks into a microprocessor-based system-on-chip (μP-SoC) to shrink system footprint and associated cost [16], [22], [23]. However, such integration often incurs silicon area increase since most of analog, mixed-signal, and digital circuits require substantial amounts of hardware to enable fast, accurate, and robust operation.

An ambient temperature sensor (T-sensor) and a physically unclonable function (PUF) are two widely used components in IoT devices. The former is a critical building block for environmental monitoring; the latter is a notable security macro used for secret key generation for cryptography and chip-ID generation for authentication. However, implementing dedicated circuits for those functions requires non-negligible silicon area, especially when they are designed for high accuracy and robustness [2]–[11], [13], [18], [20]–[23].

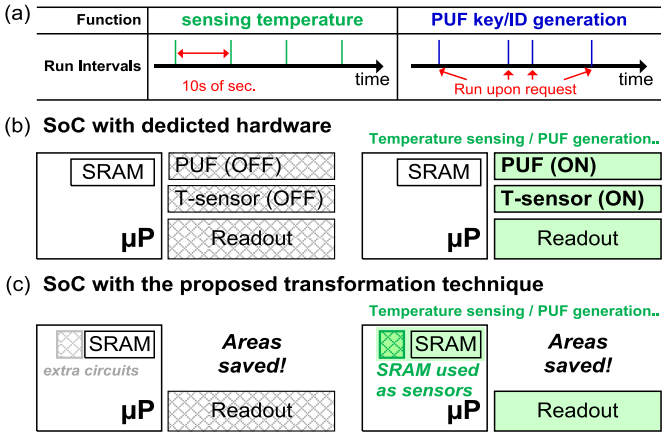


Fig. 1. (a) Low duty cycle operation such as ambient temperature sensing and PUF. (b) Dedicate hardware implementation of those functions is area inefficient. (c) Our proposed transformation approach can save hardware.

It is noteworthy that in many applications, T-sensors and PUFs exhibit low duty cycle, making the approach of dedicated hardware further inefficient in area. As shown in Fig. 1(a), for example, a T-sensor can only be active every several seconds (or even longer) since ambient temperature changes rather slowly [16], [22]. A PUF also needs to be active only upon a request for, e.g., encrypting and decrypting messages, and chip authentication processes [19], [20]. Therefore, as shown in Fig. 1(b), dedicated hardware can be idle for most of the time.

In this paper, therefore, we aim to address such area inefficiency, and propose a novel technique to transform the existing SRAM in the instruction cache (I$) of a μP into a T-sensor or a PUF [Fig. 1(c)]. This hardware recycling approach can reduce silicon footprint while integrating more features on a chip. To enable such transformation, we made a minimal amount of change in the SRAM circuits, instruction set architecture (ISA), and pipeline control logic. The outputs of the transformed T-sensor and PUF operations are stored in the data memory of the μP for post-digital processing.

Manuscript received July 20, 2017; revised October 4, 2017 and December 26, 2017; accepted December 27, 2017. This work was supported in part by the NSF under Grant CCF-1453142 and in part by the Catalyst Foundation. This paper was approved by Guest Editor Rikky Muller. *(Corresponding author: Jiangyi Li.)*

The authors are with the Department of Electrical Engineering, Columbia University, New York City, NY 10027-6902 USA (e-mail: jl3920@columbia.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSSC.2018.2791460

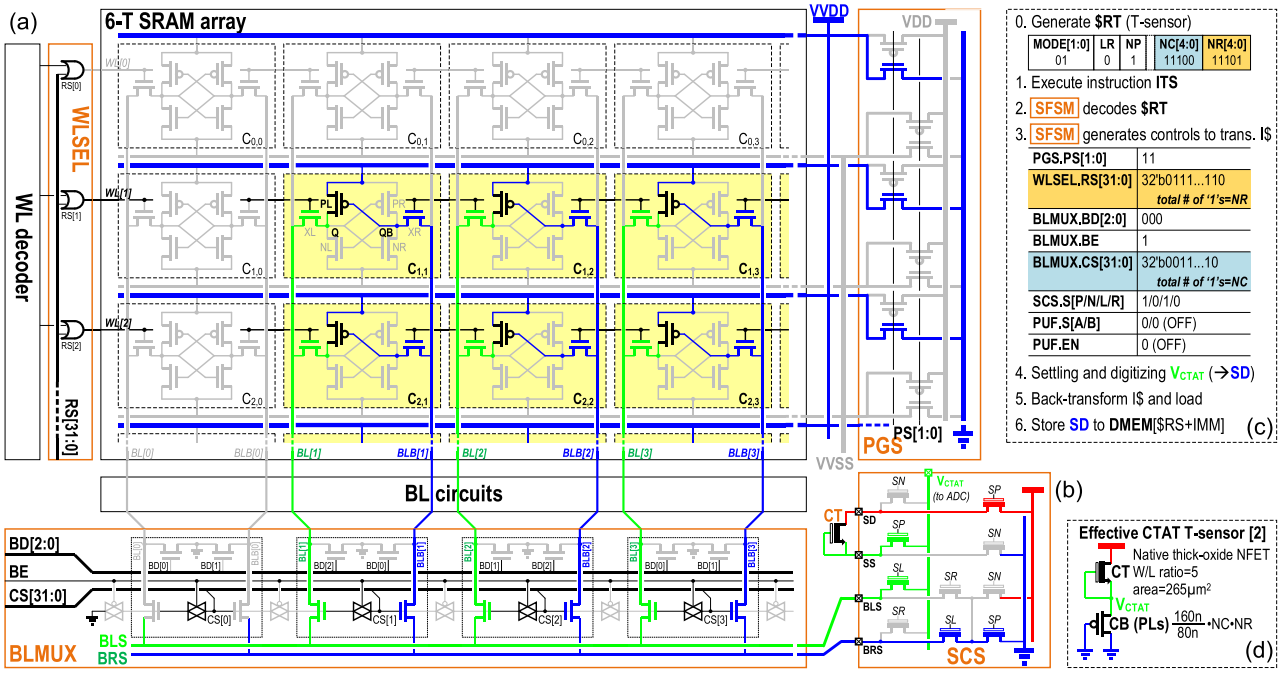


Fig. 2. (a) SRAM with the added peripherals showing the configuration for T-sensor transformation. (b) Schematics of SCS and CT. (c) $RT and control signal values. (d) Effective circuits of the transformed T-sensor.

We prototyped a μP-based SoC with the proposed technique in a 65-nm general-purpose CMOS. The μP can operate at 320 MHz at 1-V supply voltage (V_{DD}) and consumes 10.6 pJ/cycle. The transformed T-sensor achieves an error of -0.5 °C/$+1.5$ °C after one-temperature-point calibration (OPC) across 26 instances. It achieves low V_{DD} sensitivity, exhibiting only 0.46 °C error for 100-mV V_{DD} variation from 1 to 0.5 V. The transformed PUF also achieves a desirable randomness: the analog differential output shows a normal distribution with $\mu = -1.3$ mV and $\sigma = 31.2$ mV; the digitized bitstream passes all the applicable NIST tests and achieves 0.502 inter-PUF fractional hamming distance (FHD). It also achieves robustness comparable to the state of art: 0.027% unstable bit ratio and 1.97×10^{-5} bit error ratio (BER) after temporal majority voting (TMV11) and comparator (CMP) input swapping (CIS) based masking.

The proposed transformation capability increases the area of the baseline μP by 12.9% (9.2% only for the T-sensor and 9.1% only for the PUF). The first 6.3% is for the update in the SRAM circuits and the next 6.6% is for the microarchitecture modification. The standalone T-sensor [7] and PUF [10] circuits achieving the similar accuracy and robustness would consume more silicon area, that would be ~62.9% of the baseline μP area.

II. Circuits Design for Transformation

A. T-Sensor Transformation

The key idea in the proposed transformation is to convert SRAM bitcells and peripherals into target analog circuits by applying certain logic values on bitlines (BLs), wordlines (WLs), and other control signals. In the T-sensor transformation, the target analog circuit topology is a compact complementary-to-absolute-temperature (CTAT) voltage generator [2], [17].

To support such transformation, we updated the peripherals of an SRAM block (in I$) in mainly four ways [Fig. 2(a) and (b)].

1) We inserted a WL SELector (WLSEL) between the WL decoder and the WLs such that it can assert multiple WLs based on the control signal RS[31:0].
2) We added a BL multiplexer (BLMUX) in parallel with existing BL circuits. The BLMUX can connect multiple BL and BLB pairs into a pair of BLS and BRS (e.g., BL[30:1] to BLS, BLB[30:1] to BRS) based on the control signal CS[31:0].
3) We added power-gating switches (PGSs), which can change V_{DD} nodes of bitcells (VV_{DD}) to the ground (GND) level with control signal PS[1:0]. Note that all the bitcells in an array share a single V_{DD} and V_{SS} although we draw multiple PGSs to illustrate the physical layout.
4) We added sensor configuration switches (SCSs) and a T-sensor header (CT), where we used thick-oxide IO devices to suppress subthreshold leakage for better circuit isolation.

Fig. 2(c) summarizes the settings for those peripherals, stored in and accessed as a register $RT. The CTAT generator formed via the transformation mainly consists of two transistors: 1) parallel-connected pull-up (PU) pMOSs of selected 6-T SRAM bitcells and 2) a native device added for the transformation. The equivalent circuits are shown in Fig. 2(d). The bottom transistor CB is formed by aggregating the left PU pMOSs (PLs) of the selected bitcells. The top device CT is a native device added for the transformation. The gate voltage

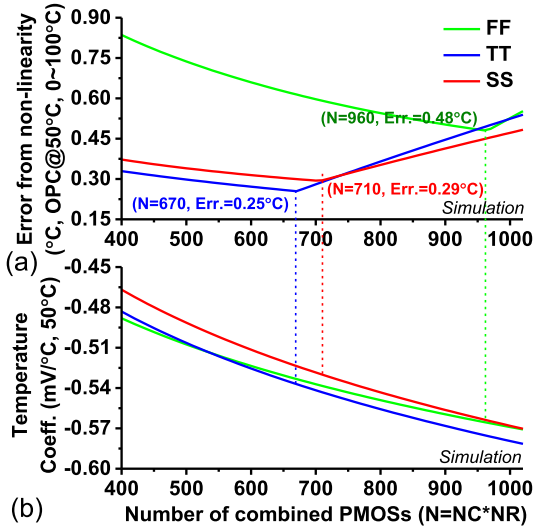


Fig. 3. (a) Accuracy-optimal NC and NR combinations across process corners. (b) Corresponding temperature coefficient.

of the CB (i.e., the BRS node from the BLMUX) is connected to GND. We can sense the output V_{CTAT} to measure ambient temperature. Note that we can use the right PU pMOSs of bitcells for the same transformation with the complimentary settings for the SCS.

Both CT and CB operate in the subthreshold region; thus, via the similar steps in [2], V_{CTAT} can be derived as

$$V_{\mathrm{CTAT}} = \underbrace{\left[\frac{k}{q} n_{\mathrm{CB}} \ln\left(\frac{\beta_{\mathrm{CT}}}{\boldsymbol{\beta_{\mathrm{CB}}}} \cdot \frac{n_{\mathrm{CT}}-1}{n_{\mathrm{CB}}-1}\right) + \left(K_{\mathrm{T1,CB}} - \frac{n_{\mathrm{CB}}}{n_{\mathrm{CT}}} K_{\mathrm{T1,CT}}\right)\right] \mathrm{T}}_{\text{temperature coefficient}} + \underbrace{V_{\mathrm{th0,CB}} - \frac{n_{\mathrm{CB}}}{n_{\mathrm{CT}}} V_{\mathrm{th0,CT}}}_{\text{offset}} \quad (1)$$

where k is Boltzmann's constant; q is the electron charge; $\beta_x = \mu_x C'_{\mathrm{ox}}(W_x/L_x)$ is the transistor strength; n_x is the subthreshold slope; μ_x is carrier mobility; C_{ox} is unit area oxide capacitance; W_x and L_x are channel width and length; K_{T1} is the temperature dependence of threshold voltage; and V_{th0} is the threshold voltage at nominal temperature.

The temperature coefficient of V_{CTAT} [the first term of (1)] is sensitive to process variations, and exhibiting non-linearity, yet can be calibrated by modulating β_{CB}, which is derived as

$$\beta_{\mathrm{CB}} = \mathrm{NR} \cdot \mathrm{NC} \cdot \beta_P = \mathrm{NR} \cdot \mathrm{NC} \cdot \mu_P C_{\mathrm{ox}}, P' \frac{W_P}{L_P}. \quad (2)$$

As shown in (2), the control signals NC and NR, which, respectively, selects the number of columns and rows of the bitcells to be combined in the transformation, can modulate β_{CB}. Fig. 3 shows the linearity-optimal NC and NR settings across process corners (simulation). The optimal NC and NR helps improve the linearity of V_{CTAT} versus temperature curves, and can be found from a batch of chips in testing. We can then perform OPC to compensate the variation of the offset which is represented as the second term of (1).

We also made several design choices to mitigate random process variation. First, we can reduce that of CB by combining a good number of PU pMOS transistors. The total transistor area of CB is $\mathrm{NR} \cdot \mathrm{NC} \cdot \mathrm{W}_P \cdot L_P$ where $W_P = 160$ nm and $L_P = 80$ nm are the PU pMOS dimensions. Moreover, to minimize the edge effects, we designed the rows and columns to be selected from the center to the edge of the bitcell array.

We made several efforts to make the sensor output V_{CTAT} robust against V_{DD} variation. Similarly from [2], we first set V_{GS} of the CT to be 0 V and increased the channel length of the CT so as to reduce the impact of drain-induced barrier lowering and channel length modulation. To digitize the outputs, an off-chip analog-to-digital converter (ADC) is used for test flexibility. It communicates with the microcontroller via a simple hand-shaking protocol.

B. PUF Transformation

Fig. 4(a) shows the circuits for the PUF transformation. The key idea is to form a pair of two-transistor threshold-voltage (V_{th})-based voltage generators [17] and compare their outputs to produce one PUF bit using a voltage CMP. This is similar to the bitcell proposed in [3].

To form a pair of voltage generators, we connect two access transistors of two adjacent bitcells (XR of $C_{1,1}$ and XL of $C_{1,2}$) to a pair of footer devices (FR and FL) in the PUF peripherals (Fig. 5) through WLSEL and BLMUX. Fig. 4(c) summarizes the control signals of the WLSEL and the BLMUX and the settings for \$RT and for PUF transformation. These make the BLMUX to connect BLB[i] to BRS and BL[$i+1$] to BLS, where i stands for the selected column index. It also pulls BL[i] and BLB[$i+1$] down to the GND level by setting two nodes, QB in the column [i] and Q in the column [$i+1$], to V_{DD} level. The two access transistors (XR of $C_{1,1}$ and XL of $C_{1,2}$) and two footers (FL and FR) now form a PUF bitcell [Fig. 4(d)].

All the devices in the effective PUF circuit operate again in the subthreshold region. Thus, the output voltage (V_{PUFL} and V_{PUFR}) can be derived as

$$V_{\mathrm{PUFL}} = -V_{\mathrm{th,XL}} + V_{\mathrm{DD}} - V_{\mathrm{th,FL}} + \phi_t \ln\left(\frac{\beta_{\mathrm{XL}}}{\beta_{\mathrm{FL}}} \cdot \frac{n_{\mathrm{XL}}-1}{n_{\mathrm{FL}}-1}\right). \quad (3)$$

The difference of the output voltages thus can be derived as

$$V_{\mathrm{PUFD}} = V_{\mathrm{PUFL}} - V_{\mathrm{PUFR}} \approx V_{\mathrm{th,XR}} - V_{\mathrm{th,XL}} \quad (4)$$

which shows that V_{PUFD} is random since it is a strong function of random V_{th} mismatch of XR and XL. Note that FR and FL are significantly larger devices and thus exhibit much less random V_{th} variations. In addition, V_{PUFD} exhibits good robustness against temperature variations since we sized the PUF footer to minimize temperature dependence, and the differential operation removes a good shared portion of the remaining temperature dependencies.

Then, by digitizing V_{PUFD} with a CMP, we can generate a PUF bit, namely, F[i]. After this, the BLMUX is configured to connect the PUF footers to the next pair of BLs (i.e., BLB[$i+1$] and BL[$i+2$]) to generate the next PUF bit (F[$i+1$]). This process continues till a PUF word (15 bits) is

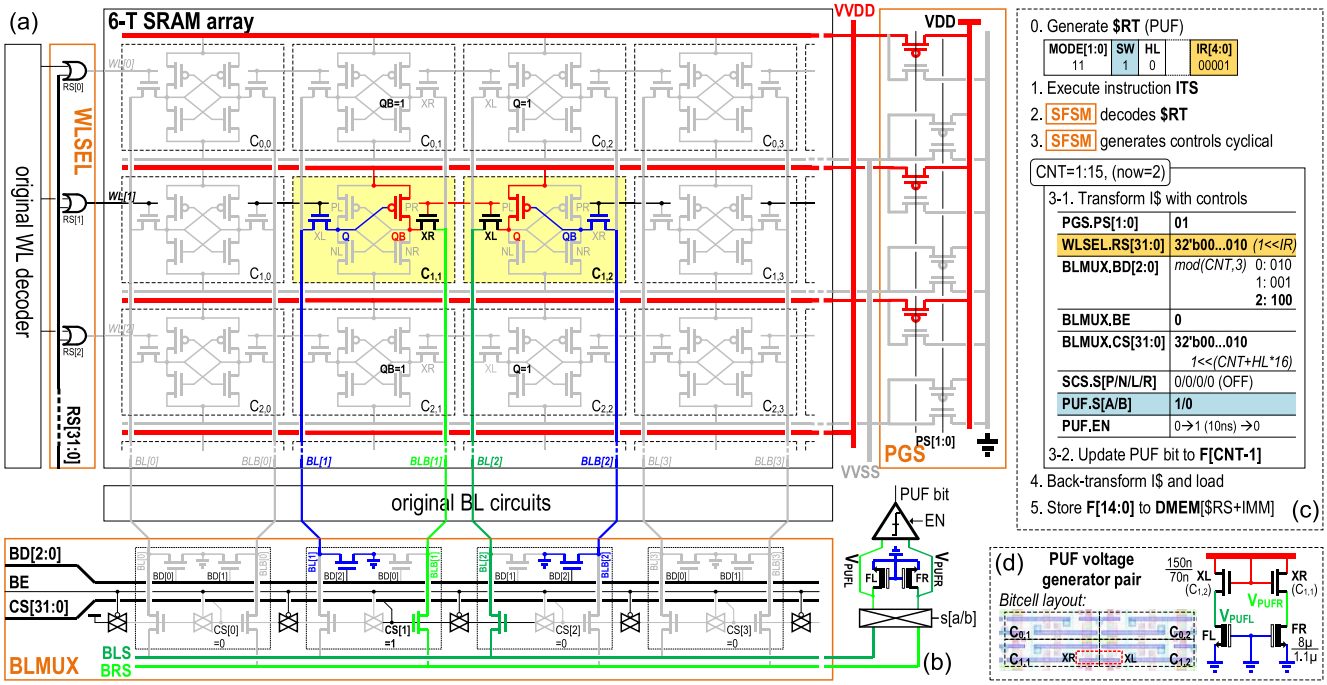


Fig. 4. (a) Circuits configurations for PUF transformation. (b) Schematics of the PUF peripherals that contains PUF footers, a CMP and an input swapper (see Fig. 5 for details). (c) $RT and control signals. (d) Effective circuits of the transformed PUF bitcell.

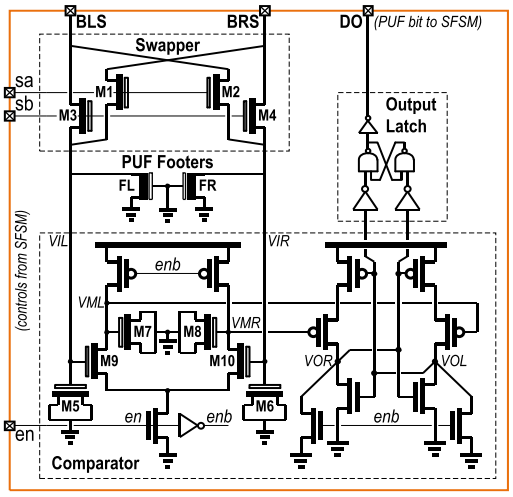


Fig. 5. Schematics of the PUF peripherals.

generated from the bitcells in the first half of the selected row in the SRAM. The process continues for the second half of the row and the other rows and produces total 960(32×30) PUF bits.

In the proposed transformation circuits, the leakage from the unselected rows could affect PUF output voltages. For example, the bitcells sharing the BLs with the target bitcells can contribute leakage to the BLs. Thus, to minimize such leakage, in the beginning of the PUF transformation, we first set all WLs high by asserting RS[31:0] and then select the target row. This ensures QBs and Qs of the unselected bitcells in the selected columns become high, creating a negative V_{GS} for the access transistors of those unselected bitcells, significantly reducing the leakage current.

The PU or pull-down (PD) transistors could also be used to form the similar structure, but we choose to use the access transistors since they undergo less transistor aging effects, allowing better bit stability over chip's lifetime [12]. In fact, negative bias temperature instability can modulate the V_{th} of PUs and PDs by as high as several tens of millivolts [15]. This makes it difficult to use them in our PUF transformation.

We also chose one access transistor from each of two adjacent bitcells since they are placed in proximity [Fig. 4(d)] and thus share the similar systematic variation. Compared to [12], which digitizes two output voltages sequentially via an off-chip ADC, this design compares V_{PUFL} and V_{PUFR} directly with a CMP. This can improve the throughput, energy consumption, and reduce quantization error.

Last but not least, we propose a method to identify unstable bitcells and generate a mask to remove them in PUF evaluation. One of the critical problems in robustness is that a PUF bitcell whose XL and XR have small V_{th} mismatch can be sensitive to noise, temperature, and V_{DD} variations. Those unstable bitcells could produce the same digital output even if we swap the inputs of the CMP, i.e., connecting BLB[i] to FR and BL[$i+1$] to FL, due to the offset of the CMP circuits, V_{th} mismatch of the PUF footers, or temporal noise.

Thus, we can perform such swapping multiple times to identify and create a mask for theses unstable bitcells. When implementing this, we leverage the configurability of the BLMUX, and therefore, it incurs little additional overhead. Note that introduction of bit-masking involves storing the mask in non-volatile memory, resulting in area overhead. The overhead can scale with the miniaturization of non-volatile memory technology [3].

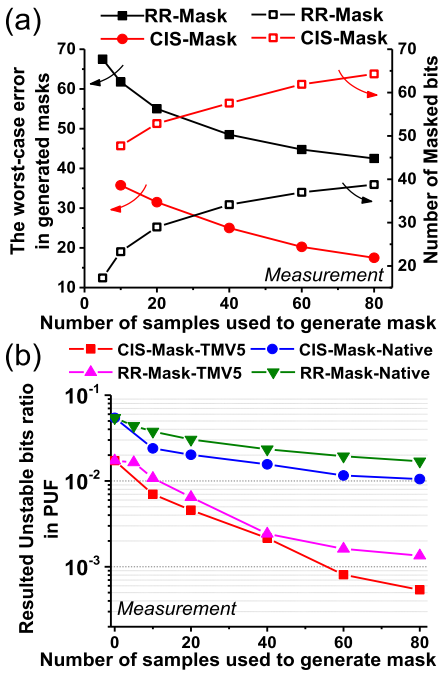


Fig. 6. (a) Accuracies of masks generated by the proposed CIS and the conventional RR techniques. (b) Unstable bit ratios post mask applications.

We compare our proposed CIS based mask generation to the conventional repetitive readout (RR) method [11], [18]. In the RR-based method, a large number (N) of repetitive PUF readings, e.g., samples are compared to identify if any bit has different readings between the samples. Similarly, in the CIS-based method, we use the total number of N samples, but with $N/2$ input-swapped and $N/2$ non-swapped. For each bit, if its readings with and without swapping are the same in any of the $N/2$ pairs, this bit is considered unstable. As shown in Fig. 6(a), the CIS method can identify more unstable bits with a less number of samples. It also exhibits the smaller worst case error. Here, the error is defined against a reference mask generated based on roughly 10$\times$ more samples (500).

We test the performance of the masks. Fig. 6(b) shows the unstable bit ratio post-mask application. The results are from the worst case mask among the total 400 masks that we generated across different sample sizes. It is shown that the CIS based method outperforms the RR method, particularly if a good number of samples are used to generate a mask. Further improvement of the mask generation can be made through body biasing the SRAM array to track the temperature dependencies of V_{th} [18]. Also, CIS and RR methods can be combined.

III. Microarchitecture Design

A. Microarchitecture Modification

To perform the abovementioned transformations and store the outputs of T-sensor and PUF to the data MEMory (DMEM), we add a new instruction called ITS [Fig. 7(a)]. This instruction has the similar format as the store word instruction in the original microprocessor without interlocked pipeline stages (MIPS) ISA [24], which stores \$RT to DMEM

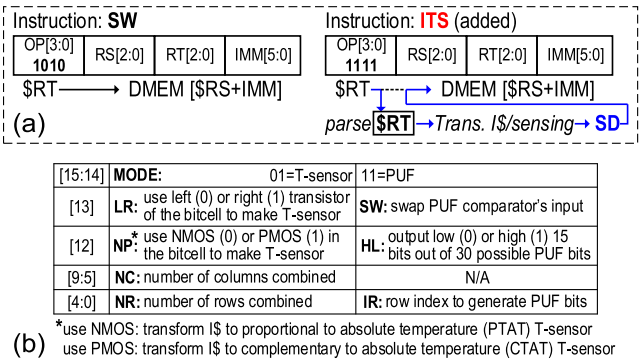


Fig. 7. (a) ITS. (b) \$RT formats for transformations.

at the address \$RS + IMM. Instead, the 16-bit register \$RT in the ITS specifies the configuration of transformation, whose value needs to be updated accordingly before the execution of ITS. The definition of the \$RT is summarized in Fig. 7(b). ITS also stores sensor output data (SD), instead of \$RT, to DMEM.

We modified the microarchitecture to support the added instruction, as is shown in Fig. 8. The baseline μP has a standard five-stage RISC pipeline and support a subset of the MIPS ISA [24]. The I\$ in the instruction fetch stage is direct mapped and can store up to 64 cache lines. The data memory of the I\$ (I\$.DATA) is made of conventional 6-T SRAM. We also updated the ID stage to support the newly added ITS and added a 2-to-1 MUX for routing \$RT. Finally, we updated the cache controller (CC) and the hazard detection unit (HDU) such that the μP handles the transformation and related operation conformal to the existing microarchitecture control.

These modifications increase the area of the μP roughly by 6.5% ($\sim$3000 μm^2). The modifications also make a moderate impact on the critical path delay, increasing it by $\sim$12%, which comes from extra logic before the WLs and extra capacitance on the BLs, although we did not optimize the critical path delay post-microarchitecture modification. On the other hand, since the SRAM bitcells are not modified, we anticipate the impact on read/write stability and noise margin is minimal. Regarding the power overhead, since most part of the added circuitry is not active during normal operation, the average power overhead is estimated $\sim$5 μW.

B. Transformation Sequence

The execution of ITS performs through four main steps.

1) It transforms the I\$ to either a T-sensor or a PUF.
2) The transformed T-sensor or PUF produces output, which is digitized and stored in DMEM.
3) It transforms the T-sensor or PUF back to the I\$.
4) The I\$ loads the next instructions to execute.

During this process, it stalls the pipeline for several cycles, which is considered as a structural hazard in the microarchitecture. The T-sensor and PUF results, stored in DMEM, can be further processed via software in the μP.

Fig. 9 describes the ITS execution in detail. First, in the $S0$ cycle, an ITS instruction enters the ID stage, asserting

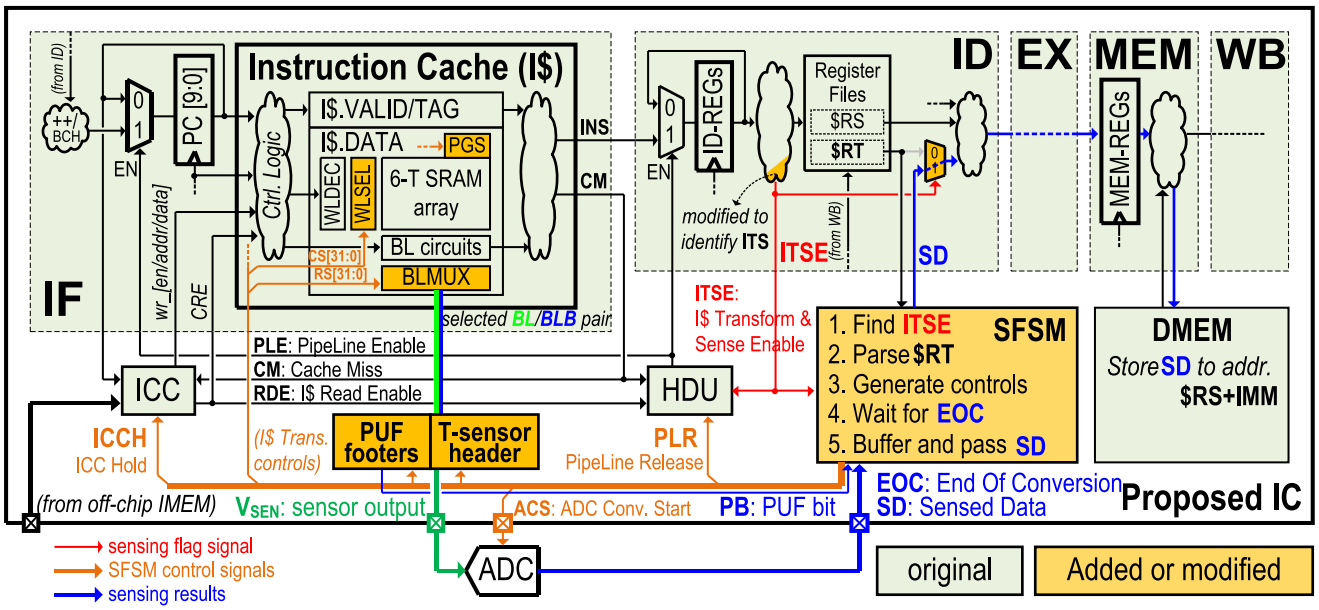


Fig. 8. Proposed μP-SoC microarchitecture. The modified and added portions are highlighted in yellow.

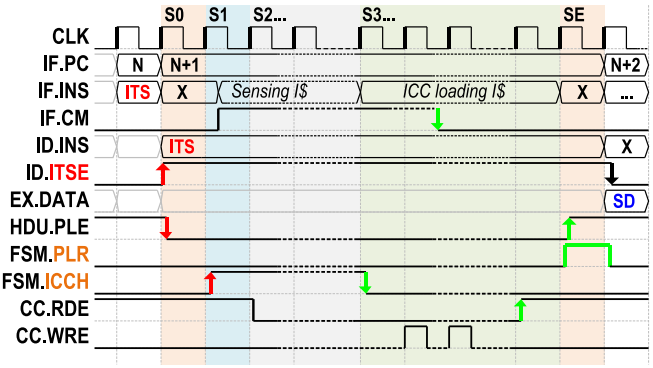


Fig. 9. Sequences of the μP-SoC of executing one ITS instruction.

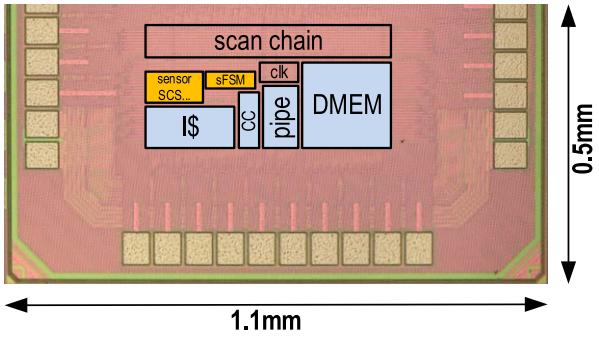


Fig. 10. Die photograph of the prototyped μP-SoC.

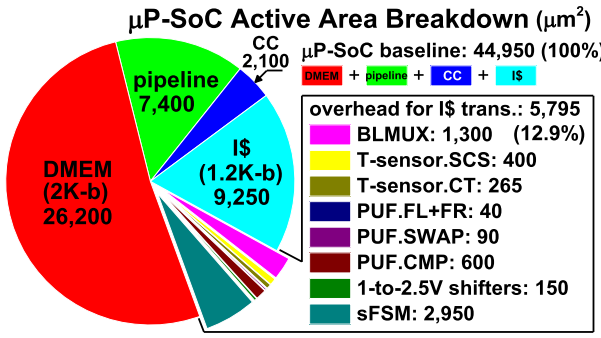


Fig. 11. Detailed area breakdown.

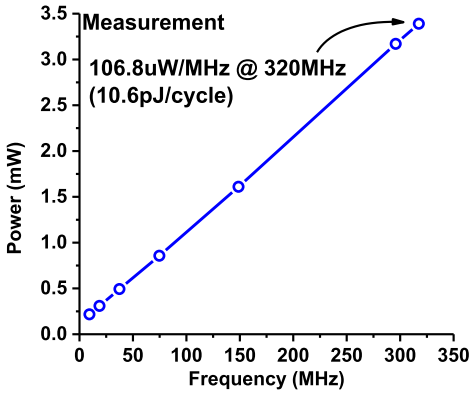


Fig. 12. Clock frequency and power dissipation of the μP-SoC.

a signal called ITSE. The HDU considers this as one of the structural hazards, stalling the pipeline by de-asserting PLE. In the next cycle ($S1$), ITSE initiates the SFSM, which decodes $RT and transforms the I$ into either a T-sensor or a PUF based on the configurations specified in $RT. Since the circuit transformation invalidates all the data stored in the I$ it asserts a cache miss flag (CM). However, the CC is notified by the cache-controller-hold (ICCH) signal from the SFSM in advance and thus ignores the asserted CM flag.

In the following cycle ($S2$), the SFSM waits for tens of micro-seconds until the analog output of the T-sensor or the

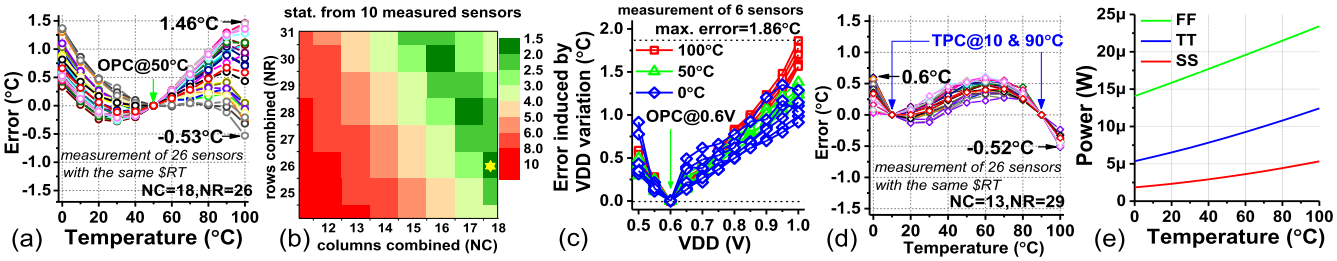


Fig. 13. T-sensor measurement results. (a) Post-OPC accuracy. (b) Post-OPC worst case error across NC and NR combinations. (c) Post-OPC accuracy across V_{DD}s. (d) Error of the transformed T-sensors after TPC. (e) Power dissipation across corners and temperatures.

NIST Test Name	Avg. P-val.	Pass?
Frequency	0.1310	YES
Block Frequency	0.7866	YES
Runs	0.9701	YES
Longest run of ones	0.0946	YES
Cumulative Sums	0.1855	YES
FFT	0.0468	YES
non-overlapping template matching	0.6270	YES
Serial	0.1299	YES
Approximate Entropy	0.1232	YES

(a)

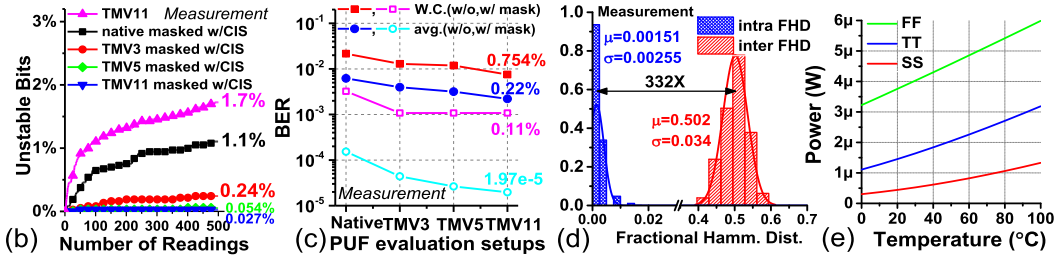


Fig. 14. PUF measurement results. (a) Applicable NIST test results on the 3712-bit PUF output. (b) Unstable bit ratios of the PUF with the TMV and CIS. (c) BER with the TMV and CIS. (d) Distributions of the inter-PUF and intra-PUF FHDs of the proposed PUF. (e) Power dissipation across corners and temperatures.

PUF settles. The output is then digitized to SDATA by an ADC (in T-sensor) or by a CMP (in PUF). Finally, in the cycle *S*3, the SFSM transforms the T-sensor or the PUF back to the I$ in $\sim$1 μs and it releases the ICCH. This makes the CC load the next set of instructions into the I$ from the main memory. Once this loading de-asserts the CM, the SFSM has the HDU to escape from the structural hazard state via a signal called pipeline release. Now the ITS instruction enters the EXecution (EX) and then MEMory (MEM) stages, where it stores SDATA in the DMEM.

IV. Chip Prototype and Measurement

Our μP-SoC with the proposed transformation capability was prototyped in general purpose a 65-nm CMOS. Fig. 10 shows the chip die photograph. Fig. 11 shows the detailed area breakdown. The additional hardware for the T-Sensor and PUF transformation takes 5795 μm^2, which is 12.9% of the 16-bit five-stage RISC μP having 1.2-kb I$ and 2-kb DMEM. If only considering sensor frontends, i.e., the modifications made in the SRAM, the area overhead is 2845 μm^2, or 6.3% of the original μP area. When counting separately for T-sensor and PUF, the area overheads are 9.2% and 9.1%, respectively.

Fig. 12 shows the performance and power dissipation measurements of the μP-SoC. At $V_{DD} = 1$ V, μ P-SoC can operate at the clock frequency as high as 320 MHz. It consumes 10.6 pJ/cycle performing a compute-intensive task (bubble sorting).

We characterized the transformed T-sensor. Across 0 °C–100 °C, it achieves the temperature sensitivity of −0.62 mV/°C. The post-OPC error is measured to be −0.53 °C/+1.46 °C across 26 instances at $V_{DD} = 0.6$ V [Fig. 13(a)] after batch trimming for NC and NR. During batch trimming, we perform OPC while sweeping NC and NR on the first ten instances and search for the optimal combination [Fig. 13(b)]. The optimal parameters are then applied to other instances. We also measured the sensitivity to V_{DD} variation across six T-sensor instances. Calibrated at 0.6 V, the T-sensors achieve a worst case error of 1.86 °C across V_{DD} variations of 0.5–1 V and across the temperature range of 0 °C–100 °C [Fig. 13(c)]. As shown in Fig. 13(d), we also tested two-temperature-point calibration (TPC) at 10 °C and 90 °C, which can reduce the error down to −0.52 °C/+0.6 °C across 26 instances. The power consumption of the T-sensor is simulated and shown in Fig. 13(e).

For the transformed PUF, we measured the randomness, uniqueness, and robustness. The differential outputs (V_{PUFD}) show a normal distribution with $\mu = -1.3$ mV and $\sigma = 31.2$ mV. The PUF bits passed all the applicable NIST random tests [Fig. 14(a)].

We performed 500 PUF bit readings and found that the unstable bit ratio is 5.39% at the nominal condition (1 V and 27 °C). We also investigate the unstable bit ratios using TMV- and CIS-based masking techniques. Fig. 14(b) shows that TMV11 can reduce the ratio down to 1.7%. The CIS-based masking technique can reduce the ratio down to 0.027% with TMV11. We tested a mask generated with 80 samples and based on the proposed CIS-based technique. It can reduce the unstable bit ratio down to 1.1% without TMV.

We also measured the BER at the nominal condition (1 V and 27 °C) across several TMV window sizes [Fig. 14(c)]. As we increase the sizes, the BER reduces. Specifically, TMV11 can scale BER down to 0.754%. The CIS-based masking (80 samples), combined with TMV11, can reduce the BER down to 0.11% and 0.00197% for the worst and the average cases, respectively; native readings 0.33% and 0.0153% BERs for the worst and the average cases, respectively.

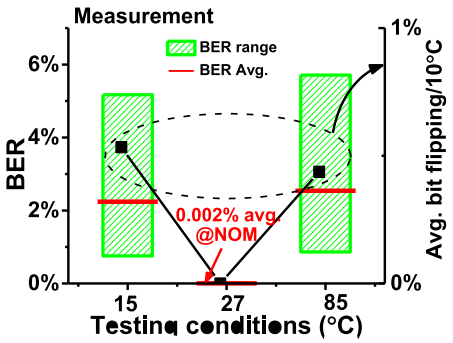


Fig. 15. BER across temperature variations.

TABLE I

COMPARISONS TO THE STATE-OF-THE-ART T-SENSORS HAVING THE SIMILAR ACCURACY AND OPERATING VOLTAGE RANGE

	Tech. (nm)	VDD range (V)	OPC error (°C)[1]	TPC error (°C)[1]	VDD sensitivity (°C/100mV)	Front-end area (μm²)[2]
This work	**65**	**0.5~1**	**-0.5/+1.5**	**-0.6/+0.5**	**0.46**	**1965**[3]
[4]	180	1.2~2	-0.5/+0.5	-	0.06	72527
[5]	180	1.2	-	-1.4/+1.5	-	48000
[6]	160	0.85~1.2	-0.2/+0.2	-	0.05	4800
[7]	**65**	1	-1.4/+1.4	-	-	2700
[8]	**65**	0.85~1.05	-	-2.3/+2.3	3.4	1180
[2]	**65**	0.6~1	-0.7/+4.7	-1.1/+2.1	0.04	400
[9]	16	0.7	-3/+3	-1/+1	-	5100
[27]	180	1.8	-	-0.1/+0.1	0.0016	40000

[1]scaled to 0~100°C [2]estimated from die photo if not report [3]BLMUX+SCS+CT

The power consumption of the PUF across temperature and process corners is simulated and shown in Fig. 14(e).

As shown in Fig. 14(d), we also characterized the uniqueness of the PUF outputs by the means of FHD. The inter-PUF FHD is measured to be 0.5017 (mean). This is close to the ideal value of 0.5, confirming that the PUF codes are highly unique. Due to the limited number of chips measured, we divided the PUF codes into four sub-codes for FHD evaluation. We also measured inter- and intra-PUF FHD distributions with TMV11 applied, which show the mean separation of 332×, exhibiting high robustness against temporal noise.

At the corner temperatures of −15 °C and 85 °C, the proposed PUF instances, respectively, exhibit the maximum BERs of 6.79% and 7.33% with TMV11. The BERs are measured by comparing the reference PUF outputs generated at the corner temperatures and at 27 °C. The temperature-induced bit-flipping ratio per 10 °C is measured to be ∼0.86%. This is calculated from the increase in average BER to rule out the impact of temporal noise. The CIS-generated mask helps reduce BERs. With a TMV11 scheme, as shown in Fig. 15, it is 0.002% at the nominal temperature, 5.28% at 15 °C and 5.82% at 85 °C. Average bit-flipping ratio per 10 °C variation is ∼0.5%. Note that we calibrated CMP's input offset voltage at 27 °C only, and thus we expect the results can be improved if automatic calibration is used across temperatures.

Finally, we compare our transformed T-Sensor and PUF to other recent works. Table I summarizes the recent state-of-the-art temperature sensor circuits that report the similar V_{DD} and temperature operating range [2], [4]–[9], [27]. One of the designs that achieves comparable performance and robustness is [7]. It achieves the post-OPC error of −1.4 °C/+1.4 °C, and its front end takes 2700 μm^2 in a 65 nm.

TABLE II

COMPARISONS TO THE STATE-OF-THE-ART WEAK PUFS ACHIEVING THE SIMILAR ROBUSTNESS AND RANDOMNESS

	Tech. (nm)	Unstable Bits %, native/post-processed		BER	Area/Bit (F²)[1]	Flip Rate per 10°C	Inter-PUF FHD	Intra-PUF FHD	Energy[4] pJ/Bit
This	**65**	**5.39/0.97**		**2.16/0.62**	**0.6k**[2]	**1.1**	**0.502**	**332x**	**0.38**[7]
[10] INV	**65**	2.34/-		-	6k	0.68[3]	0.501	140x	0.015
[10] SA	**65**	1.88/-		-	12k	0.62[3]	0.501	161x	0.163
[11]	22	30/3		8.5/0.97	9.6k	-	0.481	86x[6]	0.013
[12]	65	2.15/-		0.63/-	0.15k[5]	0.99	0.498	174x	-
[13] sym.	130	3.04/-		-	7.1k	0.68	0.506	-	0.93
[14]	65	-		4.5/-	1.1k	1.14[3]	0.491	-	-
[10] SRAM	65	16.6/-		-	0.81k	6.7	0.332	5.5x	1.1
[14] SRAM	65	-		6/-	0.19k	0.33[3]	0.497	-	-
[26] 2-Row	28	-		3.17/-	0.39k	1.3[3]	0.495	-	0.03

[1]bitcell only [2](BLMUX+PUF+level-shifters)/992 [3]estimated [4]comparing energy w/ native read [5]off-chip ADC not included [6]with dark-bits [7]simulation

Table II summarizes the comparison with recent state-of-the-art PUF circuits [10]–[14], [26]. Our proposed transformed PUF achieves substantially better robustness than the previous works based on the power-up reset states of SRAM [10], [14]. Note that [14] uses the industrial SRAM bitcells with push rules, which results in small area. However, the use of SRAM power-up states for PUF in general has low robustness. Li *et al.* [12] and Su *et al.* [13] proposed techniques to improve the robustness. Some other weak PUF circuits having the similar robustness against temperature and V_{DD} variations take silicon footprints of 25 296 μm^2 [10] to 40 374 μm^2 [11] (scaled to 65 nm) for the same code length of 928 bits. As shown in Fig. 16(a), if an SoC integrates dedicated hardware for temperature sensing and PUF, for example, [7] and [10], the area overhead is ∼27 996 μm^2, which is ∼9.8× larger than that of our proposed transformation approach (2845 μm^2). Note that the dedicated hardware approach also needs microarchitecture modification to have the interface to the T-sensor and the PUF. Thus, we consider the overhead of microarchitecture modification to be common in both approaches.

Aside from energy consumption of T-sensor or PUF operation itself, the proposed transformation technique involves reloading I$ from off-chip memory after T-sensor or PUF operation. Assuming the energy consumption for reloading I$ is 1.5 nJ per 16 bit [25] and the processor reloads all of the 64 instructions post-T-sensor or PUF operation, the energy consumption of reloading averaged over each PUF bit is 100 pJ/b (=1.5 nJ × 64/960 b). Considering this, the total energy consumption of PUF operation is 100.38 pJ/b (= 0.38 + 100). Note that the reloading energy dissipation is much larger than the energy consumption of T-sensor or PUF operation itself.

However, since T-sensor and PUF operations have low duty cycles, this energy consumption has limited impact on system power. As shown in Fig. 16(b), even if the processor performs the PUF operation every 1 ms, the average energy dissipation for PUF operation is 0.3 pJ/cycle. This is only ∼2.8% of the processor's energy dissipation per cycle (10.6 pJ/cycle).

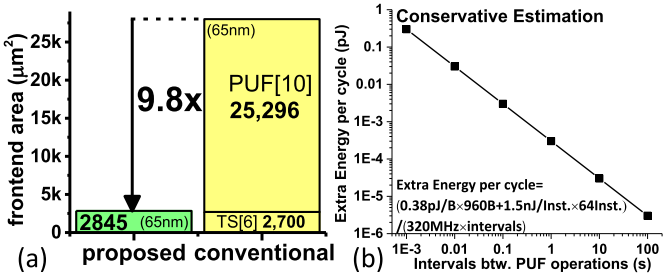


Fig. 16. (a) Area overhead comparisons. (b) Energy dissipation of PUF operation including the cache loading cost over different periods of PUF operation.

Note that here, we assume no off-chip memory access in the energy dissipation of the processor during its regular operation. Considering this, the energy impact of PUF and T-sensor operation becomes even smaller.

In addition, we also estimated the energy cost of masking and TMV in PUF operation. For masking, the extra energy cost can be roughly estimated as an extra instruction for bit-wise masking. Assuming 15 PUF bits are processed per instruction, which is the number of bits generated per PUF operation, the extra energy cost of bit-wise masking will approximately 0.71 pJ/bit. Note that the energy cost can be minimized with extra logic hardware, instead of extra instructions assumed in the above estimation. For the TMV, its energy cost can be estimated by multiplying the energy cost in native reading scheme by the number of TMV iterations. For instance, the TMV11 scheme will cost $4.18 (= 0.38 \times 11)$ pJ/bit.

V. Conclusion

In this paper, we present a μP-SoC prototype that integrates temperature sensing and PUF features for area-constraint IoT devices. We propose a transformation approach which can recycle the I$ temporarily for ambient temperature sensing or PUF code generation. The area overhead of the hardware for the transformation is 9.8× smaller than the overhead incurred by the conventional approach that integrates dedicated hardware for each feature. Measurement results of the prototyped chips show that the transformed T-sensor and the PUF achieves accuracy, robustness, and area-efficiency comparable to the state of the art.

References

[1] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Comput. Netw.*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.

[2] T. Yang, S. Kim, P. R. Kinget, and M. Seok, "Compact and supply-voltage-scalable temperature sensors for dense on-chip thermal monitoring," *IEEE J. Solid-State Circuits*, vol. 50, no. 11, pp. 2773–2785, Nov. 2015.

[3] J. Li and M. Seok, "Ultra-compact and robust physically unclonable function based on voltage-compensated proportional-to-absolute-temperature voltage generators," *IEEE J. Solid-State Circuits*, vol. 51, no. 9, pp. 2192–2202, Sep. 2016.

[4] C.-K. Wu, W.-S. Chan, and T.-H. Lin, "A 80kS/s 36 μW resistor-based temperature sensor using BGR-free SAR ADC with a unevenly-weighted resistor string in 0.18 μm CMOS," in *Proc. Symp. VLSI Circuits (VLSIC)*, Honolulu, HI, USA, Jun. 2011, pp. 222–223.

[5] S. Jeong, Z. Foo, Y. Lee, J.-Y. Sim, D. Blaauw, and D. Sylvester, "A fully-integrated 71 nW CMOS temperature sensor for low power wireless sensor nodes," *IEEE J. Solid-State Circuits*, vol. 49, no. 8, pp. 1682–1693, Aug. 2014.

[6] K. Souri, Y. Chae, F. Thus, and K. Makinwa, "12.7 A 0.85 V 600 nW all-CMOS temperature sensor with an inaccuracy of $\pm$0.4 °C (3σ) from $-$40 to 125 °C," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2014, pp. 222–223.

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

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

[9] J.-J. Horng *et al.*, "A 0.7 V resistive sensor with temperature/voltage detection function in 16 nm FinFET technologies," in *Symp. VLSI Circuits Dig. Tech. Papers*, Honolulu, HI, USA, Jun. 2014, pp. 1–2.

[10] A. B. Alvarez, W. Zhao, and M. Alioto, "Static physically unclonable functions for secure chip identification with 1.9–5.8% native bit instability at 0.6–1 V and 15 fJ/bit in 65 nm," *IEEE J. Solid-State Circuits*, vol. 51, no. 3, pp. 763–775, Mar. 2016.

[11] S. K. Mathew *et al.*, "16.2 A 0.19pJ/b PVT-variation-tolerant hybrid physically unclonable function circuit for 100% stable secure key generation in 22 nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2014, pp. 278–279.

[12] J. Li, T. Yang, and M. Seok, "A technique to transform 6T-SRAM arrays into robust analog PUF with minimal overhead," presented at the IEEE Int. Symp. Circuits Syst. (ISCAS), Baltimore, MD, USA, May 2017, pp. 1–4.

[13] Y. Su, J. Holleman, and B. P. Otis, "A digital 1.6 pJ/bit chip identification circuit using process variations," *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 69–77, Jan. 2008.

[14] R. Maes, V. Rozic, I. Verbauwhede, P. Koeberl, E. van der Sluis, and V. van der Leest, "Experimental evaluation of physically unclonable functions in 65 nm CMOS," in *Proc. ESSCIRC*, Sep. 2012, pp. 486–489.

[15] T. Yang, P. R. Kinget, and M. Seok, "Register file circuits and post-deployment framework to monitor aging effects in field," in *Proc. ESSCIRC*, Sep. 2016, pp. 425–428.

[16] M. Fojtik *et al.*, "A millimeter-scale energy-autonomous sensor system with stacked battery and solar cells," *IEEE J. Solid-State Circuits*, vol. 48, no. 3, pp. 801–813, Mar. 2013.

[17] M. Seok, G. Kim, D. Blaauw, and D. Sylvester, "A portable 2-transistor picowatt temperature-compensated voltage reference operating at 0.5 V," *IEEE J. Solid-State Circuits*, vol. 47, no. 10, pp. 2534–2545, Oct. 2012.

[18] K. Yang, Q. Dong, D. Blaauw, and D. Sylvester, "A 553F^2 2-transistor amplifier-based physically unclonable function (PUF) with 1.67% native instability," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2017, pp. 146–147.

[19] A. Van Herrewege *et al.*, "Reverse fuzzy extractors: Enabling lightweight mutual authentication for PUF-enabled RFIDs," in *Financial Cryptography and Data Security*, vol. 7397. Berlin, Germany: Springer, 2012, pp. 374–389.

[20] S. Devadas, E. Suh, S. Paral, R. Sowell, T. Ziola, and V. Khandelwal, "Design and implementation of PUF-based 'unclonable' RFID ICs for anti-counterfeiting and security applications," in *Proc. IEEE Int. Conf. RFID (RFID)*, Apr. 2008, pp. 58–64.

[21] B. Karpinskyy, Y. Lee, Y. Choi, Y. Kim, M. Kim, and S. Lee, "Physically unclonable function for secure key generation with a key error rate of 2E-38 in 45 nm smart-card chips," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Jan./Feb. 2016, pp. 158–159.

[22] M. Seok *et al.*, "The Phoenix processor: A 30pW platform for sensor applications," in *Proc. IEEE Symp. VLSI Circuits (VLSI)*, Jun. 2008, pp. 188–189.

[23] G. Chen *et al.*, "Millimeter-scale nearly perpetual sensor system with stacked battery and solar cells," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2010, pp. 288–289.

[24] *Educational 16-bit MIPS Processor*, OpenCores, Amsterdam, The Netherlands, Aug. 2013. [Online]. Available: https://opencores.org

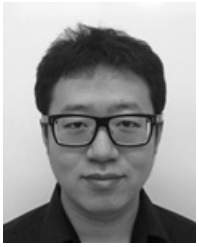
[25] M. Horowitz, "Computing's energy problem (and what we can do about it)," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2014, pp. 10–14.

[26] S. Jeloka, K. Yang, M. Orshansky, D. Sylvester, and D. Blaauw, "A sequence dependent challenge-response PUF using 28 nm SRAM 6 T bit cell," in *Proc. Symp. VLSI Circuits (VLSIC)*, Kyoto, Japan, Jun. 2017, pp. C270–C271.

[27] S. Pan, H. Jiang, and K. A. A. Makinwa, "A CMOS temperature sensor with a 49fJK^2 resolution FoM," in *Proc. Symp. VLSI Circuits (VLSIC)*, Kyoto, Japan, Jun. 2017, pp. C82–C83.

Jiangyi Li (S'17) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2012, and the M.S. degree in electrical engineering from Columbia University, New York City, NY, USA, in 2014, where he is currently pursuing the Ph.D. degree in electrical engineering with the VLSI Lab.

His current research interests include energy-harvesting systems, switched-capacitor power managements, physically unclonable function circuits, and low-power design of very large scale integration circuits.



Teng Yang (S'17) received the B.S. degree in telecommunication engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2010, and the M.S. degree in electrical engineering from Columbia University, New York City, NY, USA, in 2012, where he is currently pursuing the Ph.D. degree in electrical engineering.

His current research interests include temperature sensor designs, multimodal, and fine-grained on-chip monitoring techniques for better than worst case design and aging management technique for SRAM circuits.

Minhao Yang (S'11–M'16) received the Ph.D. degree in physics from ETH Zürich, Zürich, Switzerland, in 2015.

He is currently a Post-Doctoral Fellow with Columbia University, New York City, NY, USA, partly supported by the SNF Early Postdoc Mobility Fellowship. His current research interests include spike coding and processing, low-power spiking sensors with embedded processing, and silicon retina and cochlea.

Peter R. Kinget (M'90–SM'02–F'11) received the B.E. degree (*summa cum laude*) in electrical and mechanical engineering and the Ph.D. degree (*summa cum laude* with Congratulations of the Jury) in electrical engineering from Katholieke Universiteit Leuven, Leuven, Belgium, in 1990 and 1996, respectively.

From 1991 to 1995, he was a Research Assistant with the ESAT-MICAS Laboratory, Katholieke Universiteit Leuven. From 1996 to 1999, he was a Technical Staff Member with the Design Principles Department, Bell Laboratories, Lucent Technologies, Murray Hill, NJ, USA. From 1999 to 2002, he held various technical and management positions in IC design and development at Broadcom, CeLight, and MultiLink. In 2002, he joined Columbia University, New York City, NY, USA, where he is currently the B. J. Lechner Professor and the Chair of the Department of Electrical Engineering. From 2010 to 2011, he was with the Université catholique de Louvain, Louvain-la-Neuve, Belgium, on sabbatical leave. He is an expert on a patent litigation and a technical consultant to industry. He has authored or co-authored papers in circuits and systems journals and conferences, and co-authored three books. He holds 28 U.S. patents with several applications under review. His current research interests include analog, RF and power integrated circuits and the applications they enable in communications, sensing, and power managements.

Dr. Kinget received the Graduate Fellowship from the Belgian National Fund for Scientific Research from 1991 to 1995. He was a co-recipient of several awards including the "Best Student Paper Award—first place" at the 2008 IEEE Radio Frequency Integrated Circuits (RFIC) Symposium, the "First Prize" in the 2009 Vodafone Americas Foundation Wireless Innovation Challenge, the "Best Student Demo Award" at the 2011 ACM Conference on Embedded Networked Sensor Systems (ACM SenSys), the "2011 IEEE Communications Society Award for Advances in Communications" for an outstanding paper in any IEEE Communications Society publication in the past 15 years, the "First Prize" (U.S. $100) in the "2012 Interdigital Wireless Innovation Challenge (I2C)," the "Best Student Paper Award—second place" at the 2015 IEEE RFIC Symposium, the "Best Poster Award" at the 2015 IEEE CICC, and the IBM Faculty Award. His research group has received funding from the National Science Foundation, the Semiconductor Research Corporation, the Department of Energy (ARPA-E), and the Department of Defense (DARPA). It has further received in-kind and grant support from several of the major semiconductor companies. He was an IEEE Distinguished Lecturer of the Solid-State Circuits Society during 2009–2010 and 2015–2017, an elected member of the IEEE Solid-State Circuits Adcom during 2011–2013 and 2014–2016, and an Associate Editor of the IEEE JOURNAL OF SOLID-STATE CIRCUITS during 2003–2007, and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II during 2008–2009. He has served as a member for the Technical Program Committee of the IEEE Custom Integrated Circuits Conference (CICC) during 2000–2005 and 2016, the Symposium on VLSI Circuits during 2003–2006, the European Solid-State Circuits Conference during 2005–2010, and the International Solid-State Circuits Conference during 2005–2012.

Mingoo Seok (S'11–M'16) received the B.S. degree (*summa cum laude*) in electrical engineering from Seoul National University, Seoul, South Korea, in 2005, and the M.S. and Ph.D. degrees in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2007 and 2011, respectively.

He was a Technical Staff Member with Texas Instruments, Dallas, TX, USA, in 2011. He joined Columbia University, New York City, NY, USA, in 2012, where he is currently an Assistant Professor with the Department of Electrical Engineering. His current research interests include very large scale integration circuits and architectures, ultralow-power systems, machine-learning and cognitive computing, adaptive technique for process, voltages, temperature variations and transistor wearout, event-driven controls, and hybrid continuous and discrete computing.

Dr. Seok received the 1999 Distinguished Undergraduate Scholarship from the Korea Foundation for Advanced Studies, the 2005 Doctoral Fellowship from the Korea Foundation, the 2008 Rackham Pre-Doctoral Fellowship from the University of Michigan, the 2009 AMD/CICC Scholarship Award for picowatt voltage reference work, the 2009 DAC/ISSCC Design Contest for the 35-pW sensor platform design, and the 2015 NSF CAREER Award. He served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I from 2013 to 2015. He has been serving as an Associate Editor for the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION Systems since 2015.