SRAM-Based Unique Chip Identifier Techniques

Srivatsan Chellappa, *Student Member, IEEE*, and Lawrence T. Clark, *Senior Member, IEEE*

***Abstract*— Integrated circuit (IC) identification using unclonable digital fingerprints facilitates the authentication of ICs, device tracking, and cryptographic functions. In this paper, we present two hardware methods exploiting the inherent process-induced mismatch of SRAM cells. The proposed circuits improve upon those previously published by reducing the number of bits that vary from trial to trial, and can be used at times other than just IC power-up. The proposed circuits and methods are compared with the previous power-up approach using the experimental results from a 90-nm test chip. The required SRAM array periphery circuit changes allow the use of standard foundry SRAM cells and do not impact the memory access time.**

***Index Terms*— Authentication, digital signatures, hardware security, physically unclonable functions (PUFs), SRAM.**

I. Introduction

SECRET fingerprints or unique chip identifiers (IDs) are useful in a number of computing applications: 1) in communications, to establish the identity of the sender and recipient integrated circuits (ICs) [1], [2]; 2) in cryptographic algorithms [3]; 3) for device identification (serial numbers), to license software and track devices; and 4) to prevent gray market remarking. The IC fingerprint ID must be unique for each instance of logically identical devices and should be time invariant. The ID must be algorithmically unpredictable, ideally truly random and not pseudorandom, and repeatable under all process, temperature, and voltage conditions. The ID must be accessible, but secret from the outside world. Such constraints necessitate ID hardware (circuit) implementation rather than in software, the latter being vulnerable to attacks by memory dump or programming.

A. Contribution of This Paper

In this paper, we propose two improved methods using SRAM, exploiting the preferred state of each memory cell due to the inherent constituent device mismatch to extract a reliable ID. The modest SRAM peripheral circuit modifications do not impact the memory access time. Moreover, while the two methods proposed here are individually effective, we show that they can be used together to provide greater fingerprint reliability. The modifications allow the IC ID to be determined at all times and not just at IC power-up, providing the advantages of [4] and [5] with greater flexibility. Consequently, they can be integrated into any embedded memory (embedded use disallows board-level probing).

Manuscript received February 5, 2015; revised April 27, 2015; accepted May 28, 2015. This work was supported by U.S. Air Force Research Labs, Albuquerque, NM, USA.

The authors are with the Department of Electrical, Computing and Energy Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: srivatsan.chellappa@asu.edu; lawrence.clark@asu.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/TVLSI.2015.2445751

B. Nonvolatile Memory Hardware Chip Identification

The most frequently used method of device authentication relies on programming an ID or a digital signature in a nonvolatile (NV) memory block, such as fuses [6], [7], electrically erasable programmable read-only memory (EEPROM), or flash [8]. NV ID memory has the advantage that the fingerprint is not lost when the device is powered OFF, but the cost of having an NV memory both in terms of area and process added cost precludes this approach in some markets.

C. Identification Using Physical Unclonable Functions

Another method of generating unique fingerprints is to utilize the inherent process variations in devices to create physically unclonable functions (PUFs) [9]–[11]. Random variations affect circuit properties, and by constructing the circuits sensitive to those properties, their behavioral differences can be utilized as IC IDs. PUFs using wire delays, gate delays, and ring oscillator frequencies have been proposed [12], [13]. The transistor threshold voltage (V_T) is directly dependent on random dopant fluctuations (RDFs), and is truly random, and thus ideally suited for PUFs. Su *et al.* [14] proposed a V_T mismatch-based identification method using cross-coupled NOR cells. Since RDF is the predominant cause of SRAM mismatch in mature processes, relying on V_T strongly suggests using SRAMs for this function.

D. SRAM Power-Up State as an IC Identifier

SRAM is pervasive on modern ICs. The idea of using existing SRAM states during power-up as a fingerprint of the IC dates back to 2002 [4] and has thoroughly been studied [5]. In [5], commercial SRAMs were powered up numerous times to calculate a statistically repeatable known-ID that was then used to authenticate any other fingerprints generated from further power-ups. Unfortunately, for embedded use, this scheme suffers from the drawbacks that include lack of support for ICs with built-in self-test (BIST) and that the resulting nonmatching codes may have considerably less than ideal code separation. BIST is required in many designs to set redundancy at power-up, which means that the SRAM state will not be random when available to software or hardware normal usage. In addition, the power-up SRAM cell state is influenced by process variations internal to the cell but importantly, by external noise. As the SRAM array is powered up, cells operate in the subthreshold region where they are most easily influenced by noise, potentially producing different power-up states up in different trials.

E. Paper Organization

A brief rationale and an overview of published hardware IC ID approaches have comprised the introduction.

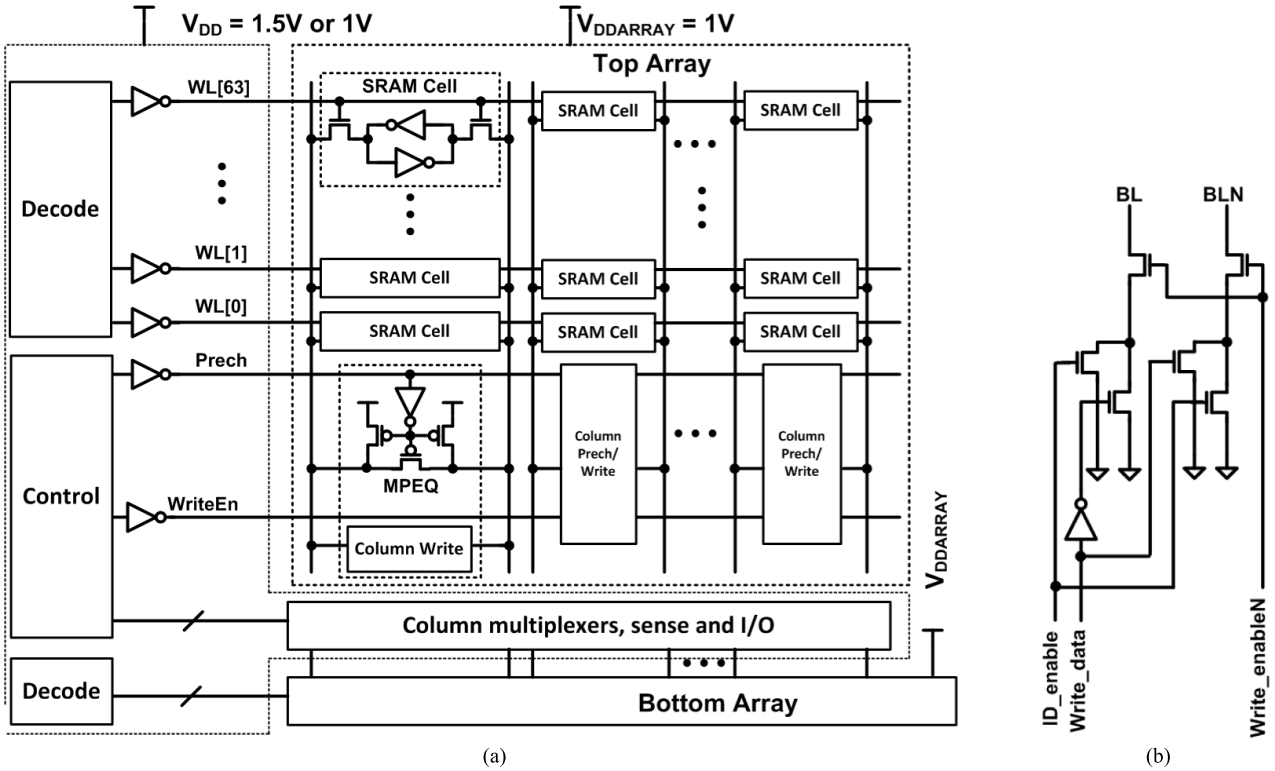

Fig. 1. (a) Part of the test SRAM showing the decoder and SRAM array with different power supply voltages used to generate the IDs. Note that the circuit becomes identical to the traditional 6-T structure when $V_{\mathrm{DDARRAY}} = V_{\mathrm{DD}}$. (b) Write circuit to implement the BL_Low method. In normal operation, the ID_enable signal is deasserted and the data and its inverse appear on BL and BLN. However, when the ID_enable signal is asserted, both BL and BLN are forced toward V_{SS}.

Section II describes the operation and circuits for the two SRAM ID schemes proposed in this paper. Section III compares the power-up SRAM ID and the proposed methods, and shows that the latter provide improved IC IDs. Analysis shows that the methods proposed here are superior to power-up by evaluating the loop gain for each case. The experimental results that verify the proposed circuits and methods are presented in Section IV. In addition, we create a more robust fingerprint by combining the proposed techniques and study the impact of temperature on the fingerprints generated from the proposed methods. Finally, the conclusion is drawn in Section V.

II. Proposed Fingerprint Generation Methods

SRAM power-up state was extensively studied as a PUF in [5]. All SRAM cells have built-in mismatch due to as-fabricated process variations. The SRAM cell is a cross-coupled inverter pair with a built-in voltage offset (V_{OFFSET}) due to RDFs, i.e., threshold voltage (V_T) and other transistor, as well as node capacitance mismatches [15], [16]. Under normal conditions, the SRAM cell's internal nodes, D and Q, shown in Fig. 2(a), are in one of two stable states DQ = 01 or DQ = 10. States DQ = 11 and 00 are unstable and thus unreachable in the normal operation. When the circuit is powered down ($V_{\mathrm{DD}} = 0$ V), the nodes D and Q are in the unstable 00 state. On power-up, the cell begins in that metastable state, but quickly transitions to one of the two stable states depending on V_{OFFSET}. Due to positive feedback in the inverter pair, any small mismatch due to process variation at the two nodes is enhanced, driving the cell to a particular favored state, similar to the sense amplifier operation. The V_{OFFSET} between the two internal nodes is a property of the SRAM cell. The preferred SRAM cell state is the one that is determined only due to the process-dependent mismatch in the SRAM cell. Cells with large process variations have a preferred state that is more noise immune and stable, while relatively well balanced cells are more likely to take on different states with each iteration.

A. Proposed Methods and Principle of Operation

In contrast to using power-up, in both the methods proposed here, we force the SRAM into a metastable state (DQ = 11 or DQ = 00), with V_{DD} applied to the SRAM cells. Thus, the SRAM state can be checked at times other than power-up, for instance, after BIST or as requested by a software application. In addition, the cell gain is lower in the superthreshold region, making the outcome more reliable for the same mismatch statistics, as described in Section III.

Meeting the SRAM cell read stability criterion ensures that internal node voltages do not increase beyond the switching threshold of the cell in the normal operation [17]. Since we wish to be able to use the SRAM in its normal configuration, i.e., for data storage and to read out the cell state after its use as a PUF, we must use the SRAM transistor sizes and layout exactly. Thus, ideally, the foundry SRAM cell is used.

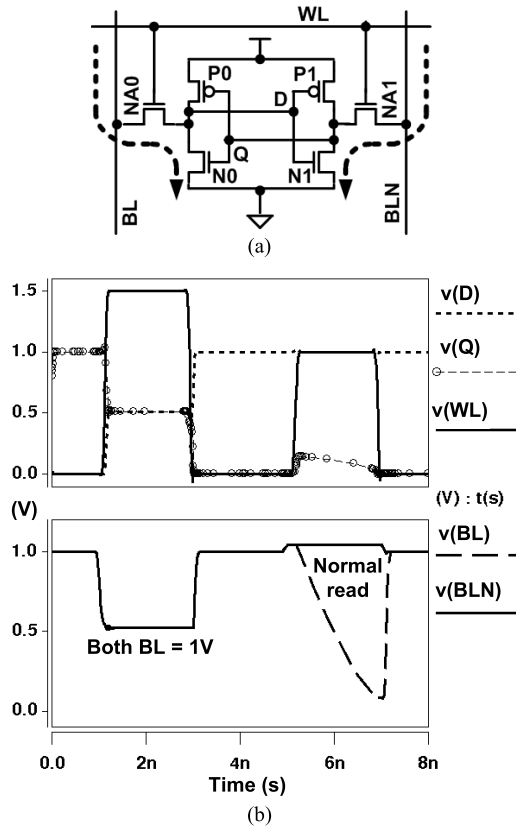

Fig. 2. (a) BL_High method drives current primarily through the nMOS access and pull-down devices NA0–N0 and NA1–NA1, respectively. (b) SRAM cell internal node (D and Q) waveforms applying the proposed method with both BLs driven to 1 V—BL voltages below show that the BLs cannot reach 1 V due to the strong nMOS pull-down transistors inside the cell. The WL is driven to a higher voltage (1.5 V here) to destabilize the cell and to 1 V, the nominal V_{DD} (1 V here) to read out the value.

The overall operation of a word of SRAM as a PUF is similar to a sense amplifier, whereby the small voltage difference due to V_{OFFSET} is amplified when the cross-coupled inverters are freed after the SRAM cell is driven to a metastable state. To force the nodes D and Q into the metastable state close to 11, the cell must be destabilized. To accomplish this, the access nMOS transistors are strengthened with respect to the pull-down transistors. This is accomplished by altering the voltages at the array level. Thus, the access transistors are made stronger than the pull-down nMOS transistors by increasing their gate overdrive, i.e., setting the word-line (WL) voltage V_{WL} above the array supply voltage $V_{DDARRAY}$ when the fingerprint is taken [see Fig. 1(a)].

Since the high V_{WL} destabilizes an entire SRAM row, that is, the smallest unit that can be driven metastable at a time. The normal readout mode is still used to determine the value of the SRAM words via normal read operations—our design has 8:1 read multiplexers above and below the sense amplifiers. The cells must be read stable—using the foundry cells provide high density and guarantee read stability. The ID capability is provided with a small modification of the SRAM memory peripheral circuits. An SRAM array provides a large number of bits to work with and ID sizes can be changed, if needed, depending upon the application.

The two proposed methods are differentiated by the voltage applied to the bit lines BL and BLN. For the first method, BL_High, i.e., both BL and BLN are driven toward V_{DD} [18], while in the second method, BL_Low, i.e., they are driven to logic 0. Note that the latter does not require higher WL voltage, although we use it here throughout.

B. Method BL_High (BLs = 1)

For normal applications, the SRAM is read or written by driving the SRAM row WL to $V_{DD} = V_{DDARRAY}$. To implement the BL_High method, the timing and control circuits are modified to allow the BLs to be precharged, while $V_{WL} = V_{DD} > V_{DDARRAY}$ destabilizes the cell. The higher voltage on the WL acts to compensate the V_{TN} drop across the nMOS access transistors when the BLs are at V_{DD}. Consequently, the minimum V_{WL} required to destabilize the SRAM cell is

$$V_{WL} \geq V_{DDARRAY} + V_{TN}. \quad (1)$$

Note, however, that the strong nMOS pull-down transistors force a voltage division, so the cell internal nodes are driven closer to $V_{DDARRAY}/2$ than $V_{DDARRAY}$. The equalization device, transistor MPEQ in Fig. 1(a) helps drive identical BL, BLN voltages, i.e., mitigate mismatch impact from the column precharge transistors. Referring to Fig. 2(a), with both BLs driven high and $V_{WL} > V_{DDARRAY}$, the SRAM cell internal node voltages V_D and V_Q are determined by the strength ratios of the access NA0 to internal transistors N0 and NA1–N1. When the cell is released from this metastable state, the small initial mismatch is amplified by a positive feedback in the coupled inverter pair and the cell transitions to one of the two stable states. Simulations of SRAM cell internal nodes driven to this high metastable state are shown in Fig. 2(b) for $V_{WL} = 1.5$ V with $V_{DDARRAY} = 1$ V. The second read, with the $V_{WL} = V_{DDARRAY} = 1$ V, is the normal SRAM read to determine the result.

C. Method BL_Low (BLs = 0)

In this proposed method, the BLs are driven toward 0 V by simultaneously writing a logic 0 to each BL. In this scheme, the dominant ratio is that between the pull-up pMOS and the access nMOS transistors, e.g., P0 and NA0, respectively, as shown in Fig. 3(a). In this proposed method, the SRAM cell internal nodes are forced to metastable voltages close, but slightly greater than the 00 power-up point. The SRAM cell is easily destabilized even without the higher voltage on the WL. Therefore, the greater than $V_{DDARRAY}$ V_{WL} voltage is not required. To keep the design common for both the proposed methods, V_{WL} is greater than V_{DDCELL} throughout this paper (this has a minor impact on the ratios, but not the mismatch). Thus, to implement the BL_Low method, the only additional design change required is to simultaneously drive the BL and BLN nodes to be equal to 0 V. This is achieved by modifying the write circuitry in the SRAM, as shown in Fig. 1(b), with an additional signal ID_Enable. When ID_Enable is asserted, the cell is written with BL = BLN = 0. The SRAM cell completes the operation to reveal the preferred state when

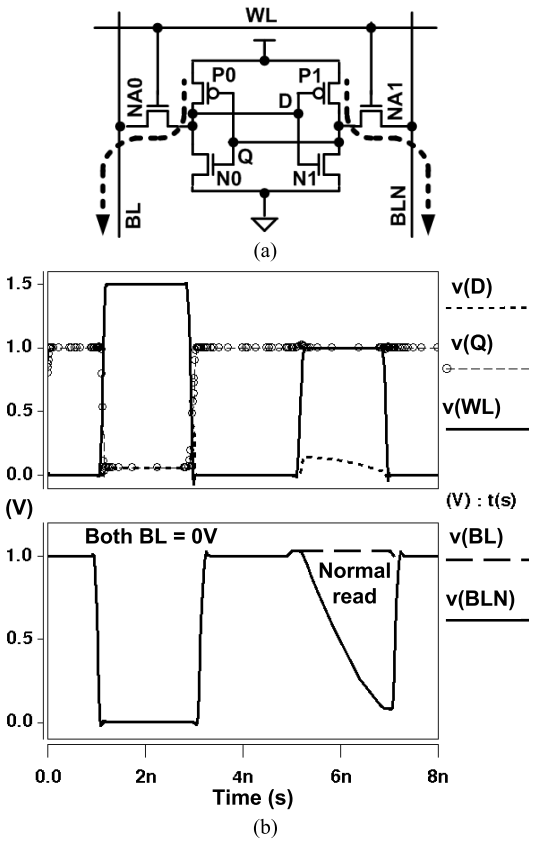

Fig. 3. (a) BL_Low method drives current primarily through the nMOS access and pull-up pMOS devices NA0 from P0 and NA1 from P1, respectively. (b) Waveforms applying the proposed method with both BLs = 0 V—when driven metastable, the BLs near 0 V, since the pMOS transistors must be weak to ensure normal write-ability.

WL is deasserted [Fig. 3(b)]. Subsequent read operations (reads from each column of the row) reveal the preferred cell states for use in the ID. One such read operation is shown in Fig. 3(b).

D. Circuit Operation

Although in both methods, the SRAM cell is forced to a metastable state, the bit-line voltage amplitude plays a significant role in determining the mismatch in the internal nodes' $V_D - V_Q$ (offset) voltage. The SRAM circuits in Figs. 2(a) and 3(a) illustrate the primary current flow through the access transistor NA0 and NA1 that creates the different voltages at nodes D and Q under different BL conditions, projecting the mismatch onto the SRAM cell logical state when the WL is deasserted.

The relative offset determined by simulating with single V_T variations is shown in Fig. 4. For the BL_High method, ΔV_T of the pull-up pMOS transistor minimally affects the SRAM offset voltage over the relevant range. Thus, this method should be relatively immune to negative bias temperature instability (NBTI) effects that primarily shift the V_T of pMOS transistors in deep submicrometer technologies [19]. Conversely, the P0 V_T shift dominates the BL_Low method, as expected. Fig. 5 shows the static noise margin (SNM) butterfly curves of the SRAM cell during the normal

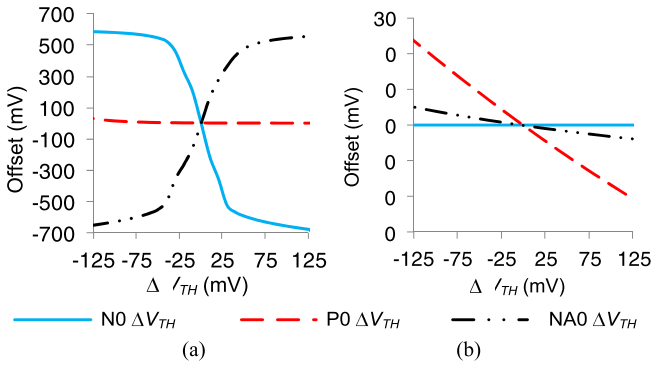

Fig. 4. (a) SRAM V_{OFFSET} sensitivity curves as a function of SRAM transistors' ΔV_{TH} for the method BL_High. (b) SRAM offset versus threshold voltage variation in SRAM transistors for the method BL_Low. In (a), ΔV_{TH} of the pull-up pMOS transistors has minimal impact on V_{OFFSET}. In (b), ΔV_{TH} of the pull-down nMOS transistors has negligible impact on V_{OFFSET}.

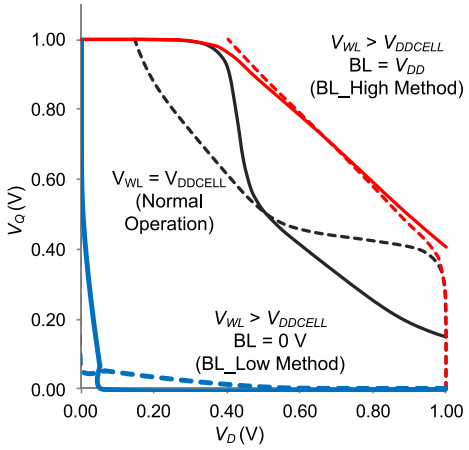

Fig. 5. SNM analysis of the SRAM at the nominal $V_{\mathrm{WL}} = 1$ V, at $V_{\mathrm{WL}} = 1.5$ V with BLs at V_{DD} and at $V_{\mathrm{WL}} = 1.5$ V with BLs at 0 V. Note that in both the proposed methods, the SRAM cell has no significant eye in the butterfly curve—BL_High metastable point is highly dependent on the BL drive.

memory operation and when destabilized by the proposed methods.

E. Power Dissipation

The proposed methods generate IDs from a single row of SRAM cells by selecting a particular WL. This minimizes the short-circuit current in the SRAM array when the cells move from a metastable state to a stable state, thereby reducing the overall power consumption. Modern SRAM cell layouts do not allow row-by-row power gating, so for the power-up scheme, the choice becomes to destabilize the entire array or by columns. In our proposed methods, a row of the array is destabilized and then can be read out conventionally by eight read operations (due to the 8:1 column multiplexing).

The simulation results show that the short-circuit current during destabilization is limited in all the cases by the SRAM cells, keeping it below 30 μA per column. This is similar to the nominal SRAM read current. The overall power dissipation of destabilizing one 256-bit row and reading it out is thus equivalent to eight 32-bit write and eight read operations.

III. Analysis

A. Noise Impact on the SRAM Fingerprint Methods

In this section, we show analytically that our proposed methods of forcing the internal D and Q nodes to 11 or 00 using a WL voltage higher than V_{DD} and then releasing them is more immune to noise and temperature fluctuations than SRAM power-up. For an inverter in the transition region above threshold, both the pMOS and the nMOS are in the saturation region. Hence, the input and output voltages V_I and V_O, respectively, are related by

$$1/2K_N(V_I - V_{\mathrm{TN}})^2(1+\lambda_N V_O) = 1/2K_P(V_{\mathrm{DD}} - V_I + V_{\mathrm{TP}})^2(1+\lambda_P(V_{\mathrm{DD}} - V_O)). \quad (2)$$

For simplicity $K_N = K_P = K$, $\lambda_N = \lambda_P = \lambda$, and $V_{\mathrm{TN}} = -V_{\mathrm{TP}} = V_T$ is assumed. With these simplifications, we obtain

$$1/2K(V_I - V_T)^2(1+\lambda V_O) = 1/2K(V_{\mathrm{DD}} - V_I + V_T)^2(1+\lambda(V_{\mathrm{DD}} - V_O)). \quad (3)$$

Differentiating both sides with respect to V_I provides

$$K(V_I - V_T)(1+\lambda V_O) + 1/2K(V_I - V_T)^2\lambda\frac{dV_o}{dV_I} = -K(V_{\mathrm{DD}} - V_I - V_T)(1+\lambda(V_{\mathrm{DD}} - V_O)) - 1/2K(V_{\mathrm{DD}} - V_I + V_T)^2\lambda\frac{dV_o}{dV_I}. \quad (4)$$

Rearranging (4) gives the inverter gain

$$\frac{dV_O}{dV_I} = -2\frac{V_I(2V_O - V_{\mathrm{DD}}) - V_T\left(\frac{2}{\lambda} + V_{\mathrm{DD}}\right) + V_{\mathrm{DD}}\left(\frac{1}{\lambda} + V_{\mathrm{DD}} - V_O\right)}{(V_I - V_T)^2 + (V_{\mathrm{DD}} - V_I - V_T)^2}. \quad (5)$$

To calculate the SRAM loop gain, we evaluate (5) for $V_I = V_D$ varying from 0 to V_{DD}. Again, for each V_D, we calculate the corresponding $V_O = V_Q$ (the cell right side voltage) using (3) and now evaluate (5) for $V_I = V_Q$ (i.e., dV_D/dV_Q) [for simplicity in our analysis, the forward and backward transfer functions are equal as the inverters are identical as per the assumption and hence we can use (3) for both cases]. The loop gain is the product of the derivatives and is shown in Fig. 6 (red dashed line).

In the power-up method, all the BL, BLN, and WL transition from 0 V to the supply voltage. Hence, the coupled inverters operate in the subthreshold region during the state resolution process. For a subthreshold inverter, V_I and V_O are related by

$$K_N\Phi_t e^{\beta(V_I - V_{\mathrm{TN}})/n}(1 - e^{-\beta(V_O)}) = K_P\Phi_t e^{\beta(V_{\mathrm{DD}} - V_I + V_{\mathrm{TP}})/n}(1 - e^{-\beta(V_{\mathrm{DD}} - V_O)}) \quad (6)$$

where $\Phi_t = kT/q$ and $\beta = 1/\Phi_t$, and n is an empirical parameter approximately equal to 1.6. Again, we assume $K_N = K_P = K$ and $V_{\mathrm{TN}} = -V_{\mathrm{TP}} = V_T$, so (6) becomes

$$K\Phi_t e^{\beta(V_I - V_T)/n}(1 - e^{-\beta(V_O)}) = K\Phi_t e^{\beta(V_{\mathrm{DD}} - V_I + V_T)/n}(1 - e^{-\beta(V_{\mathrm{DD}} - V_O)}). \quad (7)$$

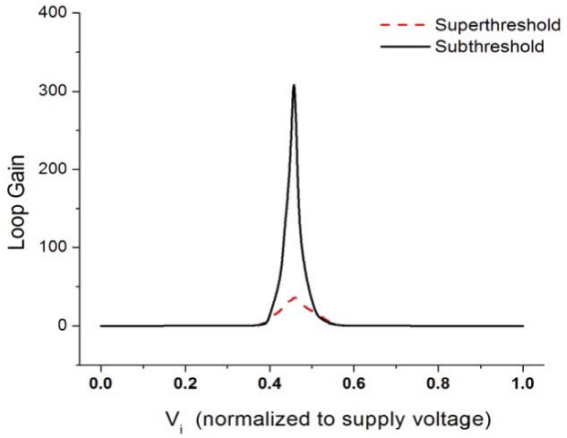

Fig. 6. Loop gain of the SRAM cell versus V_I normalized to the supply voltage.

Differentiating with respect to V_I gives

$$\frac{dV_O}{dV_I} = -\frac{e^{\beta(V_I - V_T)/n}(1 - e^{-\beta V_O}) - e^{\beta(V_{\mathrm{DD}} - V_I + V_T)/n}e^{-\beta(V_{\mathrm{DD}} - V_O)}}{n(e^{\beta(V_I - V_T)/n}e^{-\beta V_O} + e^{\beta(V_{\mathrm{DD}} - V_I + V_T)/n}e^{-\beta(V_{\mathrm{DD}} - V_O)}}. \quad (8)$$

The loop gain during power-up is determined by evaluating (8) using the same procedure as described for the superthreshold case as BL and BLN rise toward V_{DD} and calculating the product of the two inverter gains. This loop gain is also shown in Fig. 6 (solid line). It is clear that the subthreshold loop gain is much higher. The greater slope near the high gain region makes the feedback circuit significantly more prone to incorrect evaluation (away from the cell preferred state) due to noise. In addition, from (8), it is clear that the subthreshold loop gain is also highly dependent on the operating temperature. Parameters in this analysis match the foundry 90-nm process used for the experimental validation in Section IV. Regardless of the starting point, i.e., the method to destabilize the cell, any noise is amplified by the loop gain. Hence, the lower loop gain provided by destabilizing an already powered up cell should be more repeatable.

IV. Experimental Verification

A. Test Chip

An SRAM test array with 32-k cells fabricated in a bulk CMOS foundry 90-nm process was used to experimentally verify the proposed fingerprint ID methods. The test structure has separate array and WL driver supplies and allows the direct electrical measurement of the SRAM cell characteristics [20]. A micrograph of the test chip with the array layout inset is shown in Fig. 7. The test structure has the same write and sense circuits as the larger group of arrays outlined in the figure. The write circuit shown in Fig. 1(b) is optimized to alleviate crowbar currents in PUFs application, and differs from that implemented on the test chip.

B. Experiments

In the tests, 1000 trials using 1000 SRAM cells were performed using each of the methods. All three schemes,

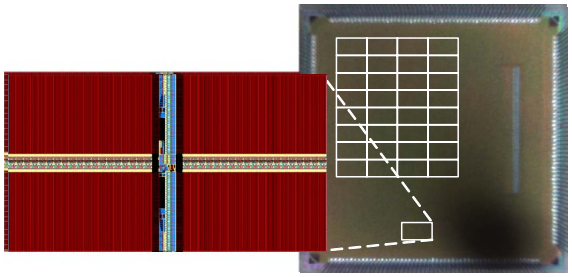

Fig. 7. SRAM test die photo and overlay of the SRAM bank test structure. In the test structure, the test array is modified to include the separate V_{DD} and $V_{DDARRAY}$ supplies, as well as ability to drive the BLs with different timing and voltages than a standard SRAM.

the power-up state, proposed BL_High, and BL_Low methods were exercised. Each of the tests was performed with the data in the SRAM cell initialized to both 1 and 0 to uncover any data remanence issues—none were found. Measured preferred state maps comprise Fig. 8(a), (c), and (e). Cells that produce a 1 state or 0 state in all 1000 trials under any given method are the cells that produce a stable (never changing) fingerprint, which are denoted by white and black colors, respectively. Those that do not produce the same state in all 1000 trials are the uncertain cells denoted by the gray color.

These uncertain cells vary from trial to trial and may cause fingerprints that include that cell to be different, even under the same environmental conditions. The uncertain cells in both the proposed methods are significantly reduced from that in the power-up method with 149 from power-up versus 19 with the BL_High and 71 via the BL_Low methods.

C. Experimental Results and Analysis

The preferred cell states were partitioned into fingerprints 32-, 64-, and 128-bit wide. The fingerprint sizes must allow margin for the uncertain bits that fluctuate. For comparison purposes, a known fingerprint (KFP) was constructed by averaging the states obtained on an odd number of trials following [5]. We used the first three trials (out of the total 1000). Cells that do not produce a stable state in the first three trials may be discarded from the sample population, though, in our experiments, none were found. The experimental methodology is described schematically in Fig. 9 for the 64-bit fingerprints. Basically, 64-bit groups are measured. Subsequent trials to the same bits should provide the same values and these IDs should not match other 64-bit groups.

A larger number of trials will be more robust, but more difficult to implement, particularly for the power-up state scheme. States from other trials are regarded as latent (unknown) fingerprints (LFPs). The KFP is the IC fingerprint that would be measured and recorded as the chip ID, and the LFPs are analogous to the fingerprints generated during authentication requests. The LFPs are quantitatively compared with the KFPs by the Hamming distance. Each mismatching bit increases the Hamming distance between a KFP and an LFP. We refer to fingerprints from the same cell groupings on every trial as matching fingerprints, which should have

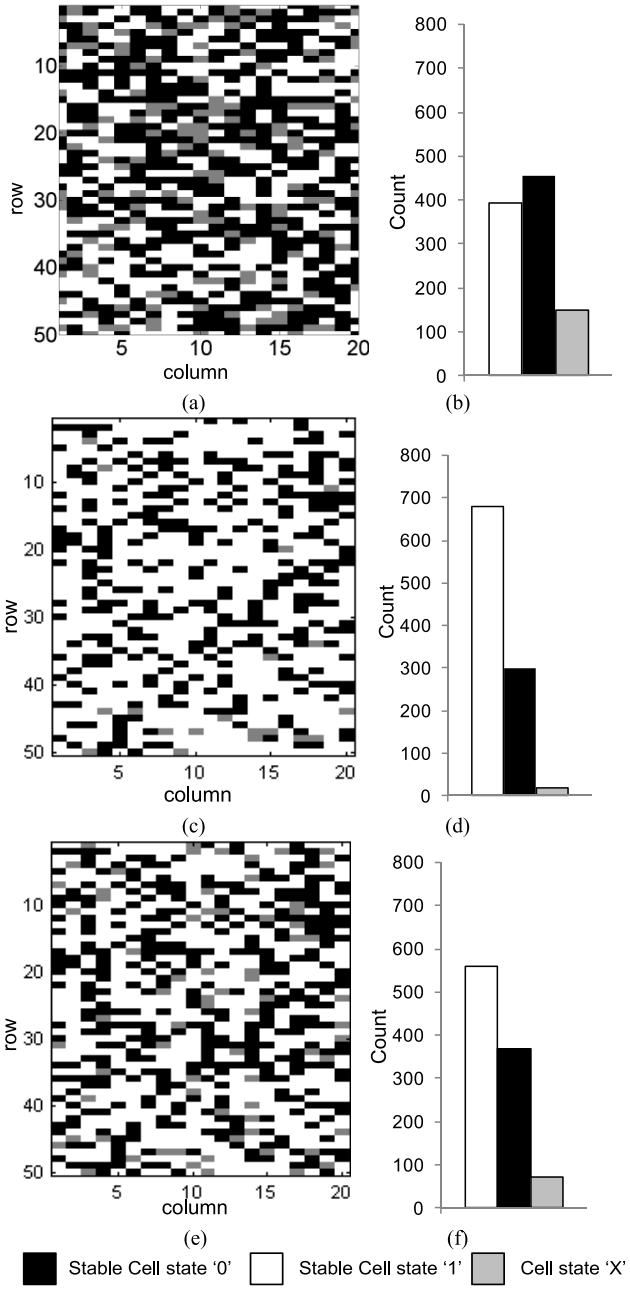

Fig. 8. (a) and (b) SRAM power-up states obtained. States obtained by testing the same cells using (c) and (d) proposed method BL_High and (e) and (f) BL_Low method. Note that the gray cells that represent uncertain states over many trials are significantly lower in the proposed methods.

low Hamming distances. Fingerprints obtained from different groups should produce large Hamming distances, i.e., should be nonmatching, as shown in Fig. 9. ID robustness requires a large Hamming distance difference between the Hamming distances of matching and nonmatching fingerprints. Any overlap between matching and nonmatching Hamming distances would result in false-positive identification or false-negative identification of ICs.

The experimentally measured results are summarized in Table I and prove the efficacy of the proposed methods for ID generation. The Hamming distances for both matching and nonmatching 64-bit fingerprints for power-up and

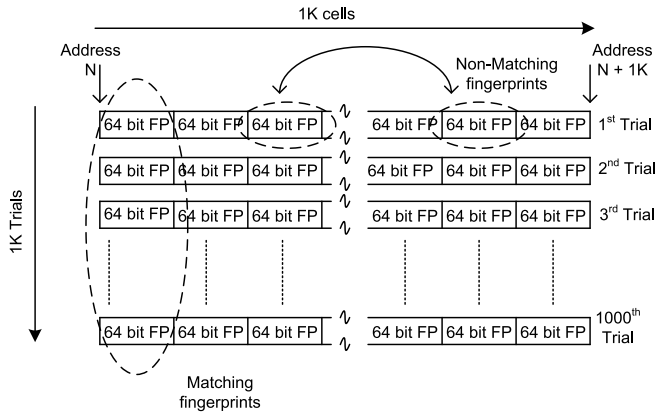

Fig. 9. Methodology for extracting matching and nonmatching fingerprints from the measured SRAM cell state data. Comparisons between matching fingerprints (the same 64-bit set) should yield low Hamming distances, while comparisons with other 64-bit groups of bits are between nonmatching fingerprints and should result in large Hamming distances. The first three trials of the same fingerprint are used for the KFP.

TABLE I
FINGERPRINT HAMMING DISTANCES OF METHODS

Method	Fingerprint Size (Bits)	Test Population (words)	Maximum Matching Fingerprint Hamming Distance	Minimum Non-Matching Fingerprint Hamming Distance
Powerup	128	7000	23	51
	64	15000	13	21
	32	31000	8	7
BL_High	128	7000	6	39
	64	15000	4	16
	32	31000	4	5
BL_Low	128	7000	10	50
	64	15000	7	18
	32	31000	5	7

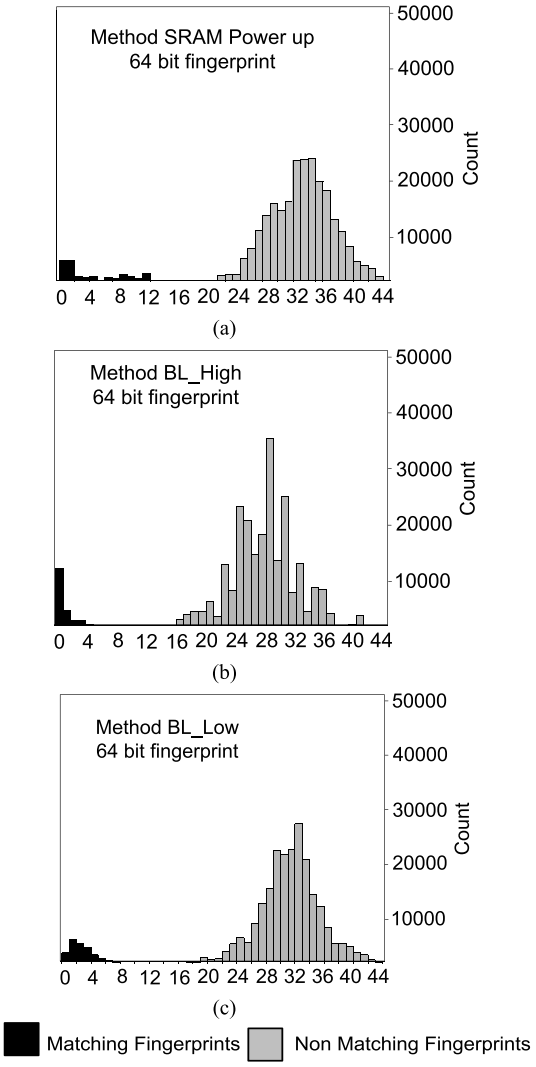

Fig. 10. 64-bit matching and nonmatching fingerprint Hamming distances of (a) power-up, (b) proposed method BL_High, and (c) proposed method BL_Low. Destabilizing the cells is evidently superior to power-up.

two proposed methods are shown in Fig. 10. In the proposed methods, the distribution of 64-bit IDs is improved with matching fingerprints having a maximum Hamming distance of four in the BL_High method and seven in the BL_Low method, as opposed to 13 when using power-up. Both the proposed methods have greater margin due to larger differences between maximum matching and minimum nonmatching Hamming distances than the power-up scheme in general. The 32-bit fingerprints in all the three methods do not provide a sufficient margin for identification, with power-up having an overlap and the other proposed methods having very small margins. Hence, 64-bit fingerprints are the smallest size considered for study in this paper, while typically larger sizes provide greater reliability.

Perfectly random preferred SRAM states would produce a mean Hamming distance of 32 for a 64-bit fingerprint given that ideally every cell has the equal probability of being 1 or 0. The power-up method owing to larger impact of noise produces a mean of 31.78. The mean nonmatching Hamming distance is 27.27 and 30.86 bits for the proposed BL_High and BL_Low methods, respectively. This observed shift is a consequence of the systematic offset with more cells preferring a 0 state or a 1 state under a given method as is evident in Fig. 8(d) and (f). Comparatively, power-up has a more even distribution of 1 and 0 states, as shown in Fig. 8(b), that causes it to have a mean nonmatching Hamming distance close to 32.

The BL_High method exhibits a significant built-in offset, evident in Fig. 8(b) as does the BL_Low scheme. A large systematic offset was also observed in [5]. Such systematic offsets can be attributed in older technologies to, among other things, polysilicon to active layer miss-registration. Modern SRAM cell layouts with single direction poly are still prone to implant induced mismatch, e.g., halo shadowing (one of the authors has measured this on 28-nm foundry SRAM cells [21]). The 90-nm cell used here has bidirectional poly, following the older layout style.

As a perfectly balanced data set with an equal distribution of 1s and 0s and no Xs, would have an ideal mean nonmatching Hamming distance of 32, it is easy to derive the expected impact of systematic skew on the mean nonmatching Hamming distance by studying their probabilities as

$$\mu_{\text{HDexpected}} = \{P_{1-0}|P_{0-1}\}\text{FP}_{\text{SIZE}} \tag{9}$$

where $\mu_{\text{HDexpected}}$ is mean nonmatching Hamming distance expected, FP_{SIZE} is the size in bits of the fingerprint, and P_{1-0} is the probability of generating a 1–0 mismatch and the P_{0-1} is the probability of generating a 0–1 mismatch under the given offset, respectively.

Therefore, an example data set with a distribution of 60% 1-states and 40% 0-states would have a mean nonmatching Hamming distance of 30.72 bits, and even the large offsets shown in Fig. 8(c) and (d) would only degrade the Hamming distance by ∼4.2 bits producing a $\mu_{\text{HDexpected}}$ of 27.8 bits (actual observed Hamming distance = 27.27 bits due to the presence of uncertain X cells). Thus, while the systematic offset in the preferred states generated is significant, the impact on the mean nonmatching Hamming distance is not large.

D. Combining the Proposed Methods

The fingerprints obtained from the two proposed methods can be used as two different and unique fingerprints. For greater reliability, the two methods can be combined to produce a more reliable fingerprint, i.e., an IC ID that has fewer uncertain cells. Such a fingerprint scheme uses some cells with states taken from the BL_High method and some with states from the BL_Low methods. XORing the results from each method will increase the number of X cells—to effectively combine the methods, one is applied and then the other is used on the cells that are uncertain from the first method. This leaves the uncertain cells as those that have insufficient mismatch in both pMOS-access nMOS and nMOS-access nMOS transistor pairs.

The values from the BL_High and BL_Low methods demonstrate this in Fig. 11(a), where the Hamming distance is again improved. Here, instead of using the first three trials to determine the KFP, we use all 1000 trials to determine any instabilities. While three trials did not show any cells X cells, this of course does as indicated by the X cells in Fig. 8. The number of uncertain (X) cells is reduced to those that are uncertain in both methods. For the 1000 addresses studied, the BL_High method produces 19 such X cells, while the BL_Low method produces 71 X cells. The combined methods produce only 5 X cells [see Fig. 11(b) and (c)]. Consequently, the combined method produces an ID with greater reliability than that from the individual methods.

E. Temperature Effects on ID Reliability

To determine the temperature stability of fingerprints generated by the proposed methods, the SRAM arrays were tested at room temperature (25 °C) and at 105 °C. The measured data from both the methods and the combined ID as proposed in Section IV-E were created. Four temperature corners were used: 1) KFP generated from data at 25 °C, while LFP was also measured at 25 °C; 2) KFP generated at 25 °C and LFP measured at 105 °C; 3) KFP generated at 105 °C and LFP measured at 25 °C; and 4) KFP generated at 105 °C and LFP also measured at 105 °C. These tests thus imitate actual possible use conditions, i.e., ID generation at one temperature but ID usage at another.

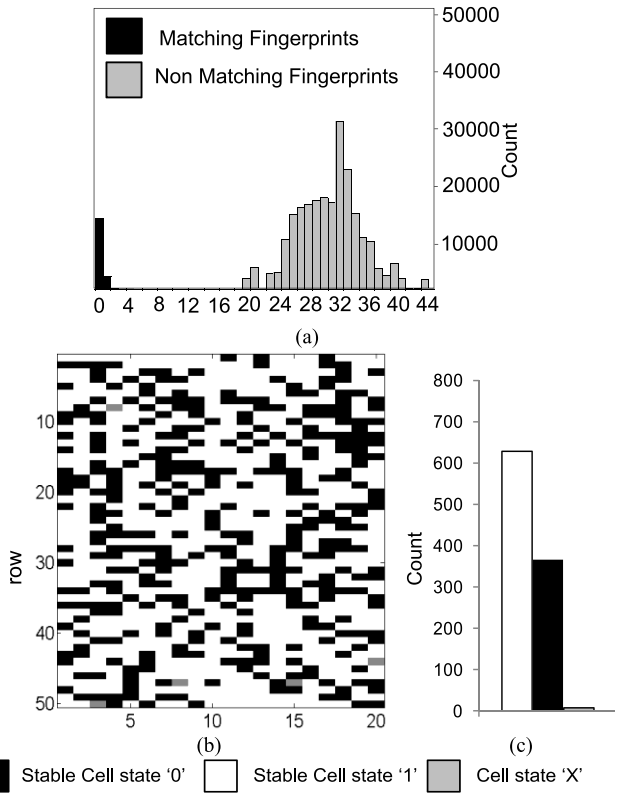

Fig. 11. (a) 64-bit fingerprint Hamming distances of the combined method. (b) Map of the cell states obtained by combining the most reliable states from the two proposed methods. (c) Number of uncertain cells has been reduced to five in 1000 trials.

The results show that the proposed methods (BL_High and BL_Low) under all temperature conditions produce reliable IC IDs. Fig. 12 comprises normal quantile plots of the nonmatching fingerprints for the proposed methods at the temperature corners. Under all temperature corners, the mean nonmatching 64-bit ID Hamming distance for all the methods are not drastically affected, being never less than 27. The worst case measured distance is at least 15 for all cases (see Fig. 12). Temperature fluctuation does increase the matching and decrease the nonmatching Hamming distances but not significantly.

The effect of temperature on the BL_Low method is more pronounced, with the matching ID Hamming distances increasing to 11 at the KFP at 105 °C, the LFP at 105 °C corner. Conversely, the BL_High method maintains matching Hamming distances less than six. The greater temperature impact on the BL_Low method than on the BL_High method is also evident from the sensitivity curves [Fig. 13(a)]. The BL_High method's greater temperature immunity is due to both nMOS transistor V_T's increasing or decreasing with temperature commensurately. In the BL_Low method, as temperature rises, the ΔV_{TH} of the pull-down nMOS transistor and access nMOS transistor both increase V_{OFFSET}, leading to greater temperature sensitivity, as shown in Fig. 13(b).

The full analysis results are shown in Table II. The combined method is the best as measured by matching Hamming distances, with measurements at the same temperature

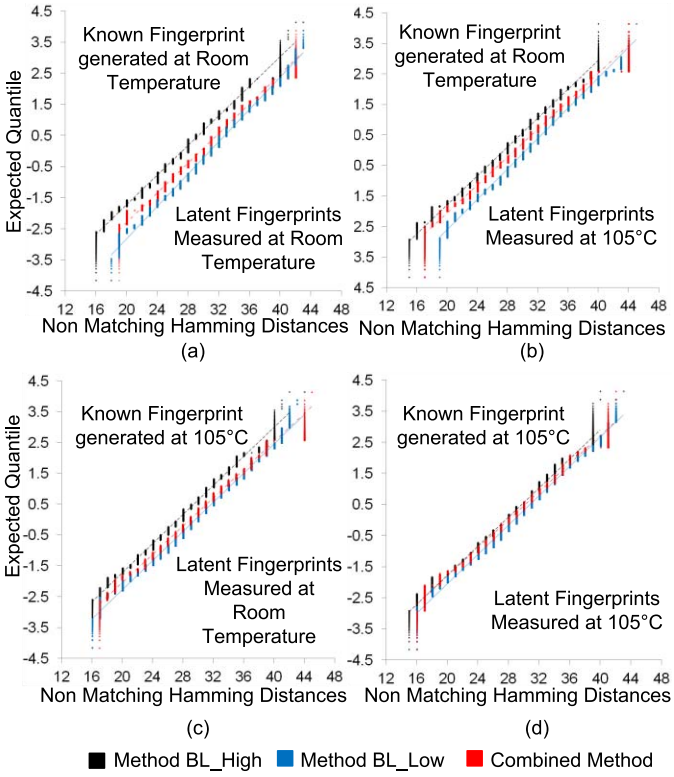

Fig. 12. Normal quantile plots of the nonmatching Hamming distances for the three proposed methods at the different temperature corners. (a) KFP at 25 °C; LFP at 25 °C. (b) KFP at 25 °C; LFP at 105 °C. (c) KFP at 105 °C; LFP at 25 °C. (d) KFP at 105 °C; LFP at 105 °C.

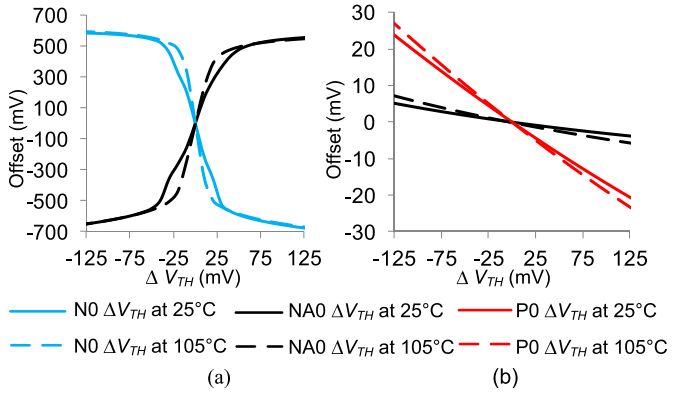

Fig. 13. (a) SRAM built-in offset V_{OFFSET} as a function of threshold voltage variation ΔV_T in the SRAM transistors for the method BL_High at room temperature 25 °C and at 105 °C. (b) SRAM V_{OFFSET} versus ΔV_{TH} of SRAM transistors for the method BL_Low at 25 °C and at 105 °C.

differing by 2 or 3 bits at room and high temperature, respectively (see Fig. 14). However, the BL_High (values shown in Table II) method is best for IDs produced at one temperature and read at another. This method produces the worst case matching Hamming distance of 5 for IDs taken at one temperature and checked at another. The worst case mismatching Hamming distance is 15 for the same case. Combining the methods, given the relatively poor BL_Low distances, does not provide a useful improvement when temperature is included in the analysis. The combined method is, however, improved over the BL_Low method.

TABLE II
HAMMING DISTANCE ANALYSIS OF 64-bit FINGERPRINTS FOR THE PROPOSED METHODS ACROSS TEMPERATURE CORNERS

	Method	KFP 25°C LFP 25°C	KFP 25°C LFP 105°C	KFP 105°C LFP 25°C	KFP 105°C LFP 105°C
Match ID Max. Ham. Dist.	BL_High	4	5	5	3
	BL_Low	7	10	8	11
	Combined	2	8	8	3
Non-Match ID Min. Ham. Dist.	BL_High	16	15	16	15
	BL_Low	18	19	16	16
	Combined	19	17	17	16
Non-Match ID Ham. Dist. Mean	BL_High	27.27	27.57	27.32	27.63
	BL_Low	30.86	30.49	29.64	28.73
	Combined	29.67	29.02	28.93	27.97
Non-Match ID Ham. Dist. Sigma	BL_High	4.16	4.19	4.21	4.25
	BL_Low	3.83	3.98	4.20	4.21
	Combined	4.17	4.33	4.35	4.39

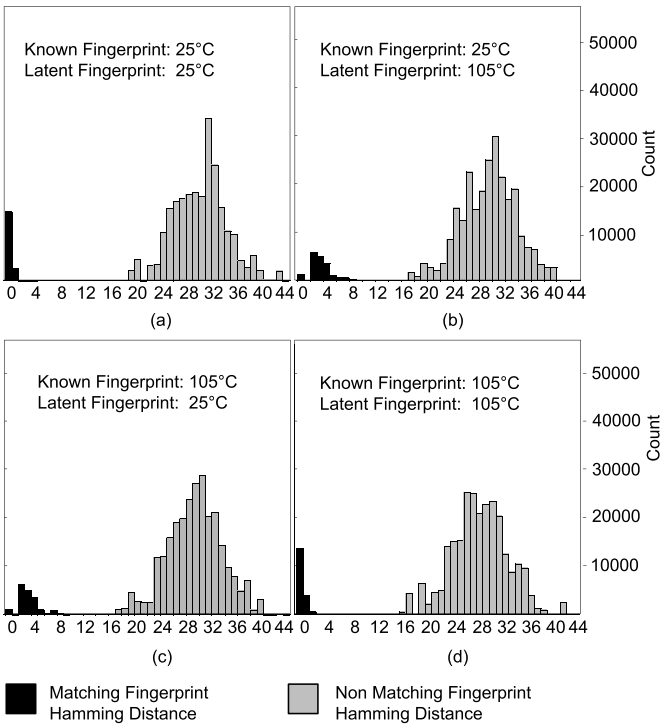

Fig. 14. Matching and nonmatching fingerprint Hamming distances generated using the combined method on the systematic offset compensated SRAM data with (a) KFP at 25 °C; LFP at 25 °C, (b) KFP at 25 °C; LFP at 105 °C, (c) KFP at 105 °C; LFP at 25 °C, and (d) KFP at 105 °C; LFP at 105 °C.

V. CONCLUSION

This paper has described two new methods for reliably extracting fingerprints (IDs) from SRAM cells based on their inherent transistor mismatch. The methods operate by destabilizing SRAM cells into metastability and releasing them to transition to their preferred state. The required circuit changes to the SRAM arrays are minor. The methods proposed are experimentally confirmed to reliably identify fingerprints of 4, 8, and 16 bytes from populations of 62 000, 30 000, and 14 000 fingerprints, respectively, on a 90-nm SRAM test chip.

The proposed techniques reduce the impact of supply variation by destabilizing powered up cells. The techniques thus

do introduce some signal noise, e.g., the WLs falling, similar to clock feedthrough. However, this noise is common mode. Moreover, WL capacitance mismatch is systematic to each cell and thus may contribute to the improved results. Both methods could be prone to precharge or write driver systematic offset if they are not large enough. This is systematic and so repeatable. We eliminated this for the weaker precharge devices, including the equalizing pMOS transistor between the BLs. If necessary, this technique is readily applied to the write driver circuit when using the BL_Low approach.

The already powered cells have considerably lower loop gain. The analysis of the loop gain of the SRAM cell shows that the proposed methods are more reliable than using SRAM power-up. Experimental results confirm the proposed methods' efficacy. The proposed SRAM ID methods require only a single SRAM row to be destabilized. They thus have low power dissipation and high speed that is similar to that in SRAM read operations. The schemes can thus easily support a large number of trials to determine cells that vary from trial to trial. The proposed methods have also been shown to work reliably at different temperatures between the ID choice and subsequent usage. The BL_Low approach is easiest to implement, as it does not require a high voltage WL. However, the BL_High method is superior, particularly due to its better performance across temperature. Moreover, this method has very low sensitivity to pMOS V_T and subsequently NBTI, which predominantly affects pMOS transistors. We believe that since the ID function will be used sporadically, powering down the ID SRAM arrays when not in use will alleviate any such wear-out issues in practical applications.

ACKNOWLEDGMENT

The authors would like to thank A. Dey for his contributions to this paper and the reviewers for catching significant errors in the initial manuscript.

REFERENCES

[1] D. Kirovski, M. Drinić, and M. Potkonjak, "Enabling trusted software integrity," in *Proc. 10th Int. Conf. ASPLOS-X*, Oct. 2002, pp. 108–120.

[2] J. Lee, D. Lim, B. Gassend, G. E. Suh, M. van Dijk, and S. Devadas, "A technique to build a secret key in integrated circuits for identification and authentication applications," in *IEEE Symp. VLSI Circuits, Dig. Tech. Papers*, Jun. 2004, pp. 176–179.

[3] *The Keyed-Hash Message Authentication Code (HMAC)*, FIPS Standard 198-1, 2002.

[4] P. A. Layman, S. Chaudhry, J. G. Norman, and J. R. Thomson, "Electronic fingerprinting of semiconductor integrated circuits," U.S. Patent 6 738 294, May 18, 2004.

[5] D. E. Holcomb, W. P. Burleson, and K. Fu, "Power-up SRAM state as an identifying fingerprint and source of true random numbers," *IEEE Trans. Comput.*, vol. 58, no. 9, pp. 1198–1210, Sep. 2009.

[6] M. Alavi *et al.*, "A PROM element based on salicide agglomeration of poly fuses in a CMOS logic process," in *IEDM Tech. Dig.*, Dec. 1997, pp. 855–858.

[7] N. Robson *et al.*, "Electrically programmable fuse (eFUSE): From memory redundancy to autonomic chips," in *Proc. IEEE CICC*, Sep. 2007, pp. 799–804.

[8] R. Glidden *et al.*, "Design of ultra-low-cost UHF RFID tags for supply chain applications," *IEEE Commun. Mag.*, vol. 42, no. 8, pp. 140–151, Aug. 2004.

[9] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in *Proc. 44th ACM/IEEE DAC*, Jun. 2007, pp. 9–14.

[10] M. Majzoobi, F. Koushanfar, and M. Potkonjak, "Lightweight secure PUFs," in *Proc. IEEE/ACM ICCAD*, Nov. 2008, pp. 670–673.

[11] T. Ignatenko, G.-J. Schrijen, B. Skoric, P. Tuyls, and F. Willems, "Estimating the secrecy-rate of physical unclonable functions with the context-tree weighting method," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2006, pp. 499–503.

[12] K. Lofstrom, W. R. Daasch, and D. Taylor, "IC identification circuit using device mismatch," in *Proc. IEEE ISSCC*, Feb. 2000, pp. 372–373.

[13] B. Gassend, D. Clarke, M. van Dijk, and S. Devadas, "Delay-based circuit authentication and applications," in *Proc. ACM Symp. Appl. Comput.*, Mar. 2003, pp. 294–301.

[14] 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.

[15] E. Seevinck, F. J. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," *IEEE J. Solid-State Circuits*, vol. 22, no. 5, pp. 748–754, Oct. 1987.

[16] A. J. Bhavnagarwala, X. Tang, and J. D. Meindl, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," *IEEE J. Solid-State Circuits*, vol. 36, no. 4, pp. 658–665, Apr. 2001.

[17] K. Agarwal and S. Nassif, "Statistical analysis of SRAM cell stability," in *Proc. 43rd ACM/IEEE DAC*, Jul. 2006, pp. 57–62.

[18] S. Chellappa, A. Dey, and L. T. Clark, "Improved circuits for microchip identification using SRAM mismatch," in *Proc. IEEE CICC*, Sep. 2011, pp. 1–4.

[19] D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *J. Appl. Phys.*, vol. 94, no. 1, pp. 1–18, Jul. 2003.

[20] S. Chellappa *et al.*, "In-situ characterization and extraction of SRAM variability," in *Proc. 47th ACM/IEEE DAC*, Jun. 2010, pp. 711–716.

[21] L. T. Clark, S. Leshner, and G. Tien, "SRAM cell optimization for low AVT transistors," in *Proc. IEEE ISLPED*, Sep. 2013, pp. 57–63.

Srivatsan Chellappa (S'15) received the bachelor's degree in electrical and electronics engineering from Anna University, Chennai, India, in 2007, and the master's degree in electrical engineering from Arizona State University, Tempe, AZ, USA, in 2009, where he is currently pursuing the Ph.D. degree.

He was with the IBM Semiconductor Research and Development Center, East Fishkill, NY, USA, as an Intern, in 2012, where he was involved in 3-D memory designs and hardware security using intrinsic IDs. He is involved in designing clock networks, memory, and sequential circuits for radiation-hardened microprocessors for use in space applications. His current research interests include hardware security, memory design, clock distribution, and radiation hardened circuits.

Lawrence T. Clark (SM'01) received the B.S. degree in computer science from Northern Arizona University, Flagstaff, AZ, USA, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from Arizona State University, Tempe, AZ, USA, in 1987 and 1992, respectively.

He was with Intel Corporation, Santa Clara, CA, USA, as a Test Engineer, and VLSI Technology Inc., San Jose, CA, USA, performing PC chipset design. In 1992, he joined Intel Corporation, where he contributed to the Pentium, Itanium, and XScale microprocessor designs, compact modeling, and CMOS imager projects. He was a Principal Engineer and the Circuit Design Manager on the XScale efforts. He joined Arizona State University (ASU), Tempe, AZ, USA, in 2004, where he hold the rank of Professor. From 2009 to 2014, he was also with SuVolta Inc. (on partial leave from ASU), Los Gatos, CA, USA, as a Chief Architect. He has authored over 110 refereed technical papers and seven book chapters, and holds over 100 patents.

Prof. Clark received the Intel Achievement Award for the record-breaking XScale performance and low power characteristics. He is a member of the IEEE Computer Society and the Association for Computing Machinery. He has been an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II, and a Guest Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I, and twice for the IEEE JOURNAL OF SOLID-STATE CIRCUITS.