Exploiting Accelerated Aging Effect for On-line Configurability and Hardware Tracking

Yang You

EECS Department
Northwestern University
Evanston, IL 60201
U9T4M8@u.northwestern.edu

Jie Gu

EECS Department
Northwestern University
Evanston, IL 60201
jgu@northwestern.edu

ABSTRACT

Conventional CMOS technology lacks an efficient way of realizing reconfigurability, which is a highly desired feature for applications such as hardware tracking for security. On the other hand, the traditional "undesirable" aging effect has presented a non-volatile "memory" to the CMOS chip. This paper exploits the aging effects in standard CMOS to enable non-volatile configurability to the chip for application of hardware tracking. A novel accelerated aging circuit is developed to shorten the required stress time to a few seconds of operation. Due to the significant challenges posed by process variation in advanced CMOS technology, a novel stochastic processing methodology is proposed to significantly reduce the failure rate of the tracking and detection. Combining both circuits and system level acceleration, the work of chip usage tracking can be realized within seconds of usage in contrast with days of operation from previously reported aging monitor. The design was implemented and simulated in 45nm CMOS technology with less than 25µW power consumption and compact sizes for easy insertion as a silicon IP using only core transistors. The robustness of the proposed stochastic processing technique has been verified using transistor level Monte-Carlo simulation. Compared with existing aging monitors, the proposed techniques accelerate the process by thousands of times enabling the desired online configurability.

Keywords
Controlled Aging, Configurability, Stochastic Processing

I. INTRODUCTION

The continuous scaling of CMOS technology allows more functionality to be integrated into a single chip. Although the vast integration brings a significant benefit of the cost reduction and performance enhancement, the need of efficient tracking and controlling the usage of a large system-on-chip (SOC) IC have become a pressing challenge. Examples of the applications that require effective tracking of the chip usage include commercial ICs that are enforced by restricted lifecycles, such as an evaluation version of the chip that needs to be disabled after a few trials; High performance ICs under export control that needs to be disabled after several usage to avoid unauthorized handling; Recycled chips that need to be detected and prevented from flowing into markets; Chips that may need user reconfiguration after its usage [1-4]. Until now, the above tasks have not been cost-efficiently realized in conventional CMOS technology because of the lack of real-time programmable non-volatile device which is alternatively offered from emerging devices such as Spintronic based SRAM or MRAM devices [5-6]. Unfortunately, such emerging devices have not been ready for mass production and likely require higher manufacturing costs. As efficient tracking circuits have become indispensable for many applications such as hardware security, this paper exploits accelerated aging effect to implement an energy and cost efficient hardware tracking circuit using conventional CMOS technology without special technology enhancement.

Silicon aging monitor has been proposed to track the usage of the standard CMOS device [7-8]. The silicon aging monitor normally consists of two ring oscillator circuits with one used as a reference and the other one to detect the stress time. Although such a circuit could provide information on the previous usage of the chip, there are a few major limitations from previously reported CMOS aging monitor circuits. First of all, the silicon monitor circuit based on natural aging is not able to reliably monitor a very short-term operation where a single usage could be finished within seconds to minutes of operation. The recent study on detection of recycled ICs shows minimum three to fifteen days of stress is needed to create a reliable detection from a ring oscillator based CMOS aging sensor due to the slow development of silicon aging effect and large process variation [9-10]. Fig. 1 shows the speed degradation of an 11-stage ring oscillator circuit under 10 seconds of stress in a cold temperature in comparison with the random process variation impact. It is seen that the ever-increasing random mismatch in advanced CMOS technology could easily overwrite the aging effect under only a few seconds of stress. To overcome the mismatch impact, a more than 8 hours continuously DC stress without recovery is required based on an optimistic analysis. Secondly, for a fresh chip, the comparison output between a reference unstressed circuit, e.g. RO and the stressed circuit is random. Therefore, a clear distinction methodology needs to be developed to obtain decisive results on the usage of the chip. In summary, the existing aging monitors do not provide the required fast and deterministic tracking results as needed.

To enhance memory creation from the CMOS technology, several technology extension has been built to enable storage and reconfiguration of the fabricated chips such as anti-fuse, floating gate device and EPROM. Anti-fuse techniques require a high voltage and high current to create an oxide breakdown, e.g. larger than 4V as reported in a 45nm technology [11-12]. Such a requirement requires high voltage device, special operating configuration and supporting circuitry to establish the programming function. Meanwhile, the use of floating gate device or other flash based memory cell will increase the manufacturing cost of the chip [13].

Recently, several physically unclonable function (PUF) circuits have been developed to create secret codes that are needed for hardware security encryption [14-15]. By utilizing

978-1-5090-1558-0/17/$31.00 ©2017 IEEE

the random mismatch of the transistors, random numbers can be created. Interestingly, aging effect has been shown to provide beneficial impact to the design. For example, a sense amplifier based random number generator is proposed featuring an offset minimization technique using hot-carrier-injection to remove mismatch of the input pairs of sense amplifiers [16]. HCI effect has also been utilized to repair read failure on bitcell in SRAM by increasing threshold voltage of the select device. It was shown that within 20 seconds, significant threshold voltage shift can be introduced to help repair the SRAM cell [17]. Different from previous application for random code generation, this paper proposes a novel latch based comparator circuit to track the chip usage. The contribution of this work is summarized below: (1). A novel aging accelerating circuit is developed to create a "tracker" of operation during real-time operation within only a few seconds of operation. For embedded usage, all the supportive transistors in the proposed design are core transistors operating without overstress. (2). A stochastic processing methodology is developed to ensure the detection accuracy with tolerance to process variation.

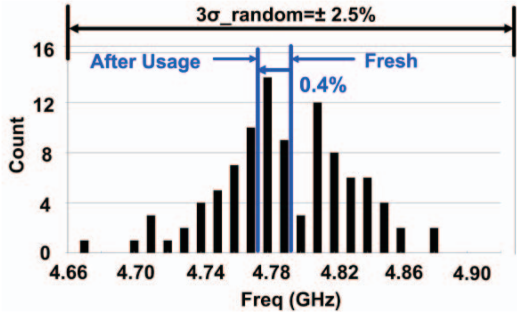


Fig. 1 The speed variation of 11-stage ring oscillator under aging degradation of 10 seconds in 0°C temperature compared with intrinsic mismatch impact in 45nm CMOS technology. Monte-Carlo simulation with 4% random threshold voltage variation is applied in the experiment. A minimum of 8 hours continuous stress is needed to overcome the mismatch impact.

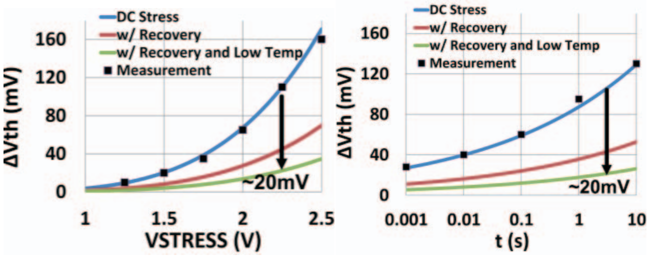
The remainder of this paper is organized as follows. Section II describes the aging model and reliability choices used in this paper. Section III describes the proposed aging accelerating cell design. Section IV describes the proposed system level design. Section V proposes a novel stochastic processing technique and associated optimization technique. Finally, conclusions are drawn in section VI.

II. CONTROLLED RELIABILITY EFFECT

To realize non-volatile storage as an operation tracker in a conventional CMOS device, the reliability degradation in modern CMOS transistors provides a potential solution. The common reliability effects include hot carrier injection (HCI), negative-bias temperature instability (NBTI), and time-dependent gate oxide breakdown (TDDB), etc. Extensive study has been performed so far to model the degradation from each of the effects. Although all three effects can be used to track the operation of the circuits, a few practical considerations are evaluated, e.g. power and cost. To create an accelerated HCI effect in a NMOS transistor within a short period of time, a substantially large current has to be delivered through the transistor leading to a large power consumption and reliability consideration of supporting circuits and wires. To enhance TDDB effect, a large on-chip voltage e.g. 4V has to be generated and safely delivered onto the device under test [11]. Such an operation requires more hardware costs and high voltage devices. Given both power and cost consideration, NBTI is used in this paper as the basic effect to create the silicon tracking circuits. However, NBTI suffers from major drawbacks of recovery as well as temperature dependency. Such drawback significantly reduces the impact of NBTI and degrades the detectability of the proposed circuits under large process variation. To overcome the deficiency of NBTI stress effect, a stochastic based design methodology is used as will be shown in section V. To correctly model the NBTI stress effect, an equation shown in (1) is used to simulate the impact of voltage, duration, temperature and recovery.

$$\Delta V_{th} = A \cdot t^{n} \cdot V_{stress}^{m} \cdot (1-\eta^{0.5}) \cdot e^{(-\frac{nE_a}{kT})} \qquad (1)$$

where A, m is a curve fitting coefficient from previous measurement data, n is power-law time exponent, η is recovery coefficient and E_a is the activation energy. The equation is calibrated based on the reported NBTI measurement in a commercial 45nm technology using the curve fitting coefficient reported in [18-19]. The recovery in equation (1) is based on previous analysis and assumes a recovery of 60% (η=0.35) [20]. The temperature dependency is also derived based on a widely used reaction-diffusion model referring to previous analysis in a 45nm technology [21-23]. Fig. 2 shows the simulation results of the NBTI effect. Considering recovery and temperature dependency, a reduction of ~5X from the initial Vth shift has been seen at the worst case condition, e.g. low temperature (0°C) and DC recovery. As a result, a target minimum stressing voltage of 2.2V and a worst case Vth shift of 20mV in a minimum of 2 seconds of stress is assumed in this work. As the aging effect is highly technology dependent, it is expected that the proposed circuits to be readjusted based on different process condition.



(a) (b)

Fig. 2 The NBTI model used in this paper in comparison of the measurement results reported previously [18-19]. (a) ΔVth vs stress voltage; (b) ΔVth vs stress time.

III. Latch-based Aging Accelerator Circuit

To enhance the aging effect on a core PMOS transistor, an elevated voltage is applied to the gate the PMOS transistor for a short period of time to induce controlled NBTI effects. Three major challenges are considered in the design of the aging accelerator circuits: (1) No transistor except the target aging transistor should experience an accelerated stress; (2) The mismatch impact of the supporting transistors should be minimal on the comparison results; (3) The circuit should be protected from aging during normal operation once the detection is finished after chip is powered up.

Fig. 3 shows the schematic of the latch based aging accelerator cell and its operating modes. Fig. 4 shows the simulated corresponding signal waveforms stress and comparison phases. During stress phase, the source voltage of MP1 is lifted to the target voltage Vddh around 2.2V while the gate is grounded (it is within the oxide breakdown limit for a short period time). Because MP1 is turned on in this configuration, the Vddh is also applied to the gate of MP2 and the NMOS device MN1-3. To avoid stress to be created, the internal virtual ground of the latch Vgnd is lifted to internal Vdd so that no elevated stress falls on any of the supporting transistors except MP1. Two transmission gates are added to the gate nodes of MP1 and MP2 to isolate the gate of MP1 from the rest of the circuits to execute stress without creating tension on other devices. To avoid stress on the transmission gates, both the body and the gate (signal Cmp') of the MN6 is also switched to Vdd during stress phase. Other control voltages such as $\overline{\text{Prech}}$, $\overline{\text{Cmp}}$ are also set to Vdd to prevent stress. For instance, although MP6 has source connected to Vddh, its gate is connected to Vdd through $\overline{\text{Prech}}$ and thus there is no high voltage across any two terminals of the device. Essentially, the raise of virtual ground and use of Vdd for control signals shift the voltage-level up by 1.1V without introducing high voltage across any two terminals of all the transistors except MP1. As a result, all transistors are well protected without overstress between any two terminals.

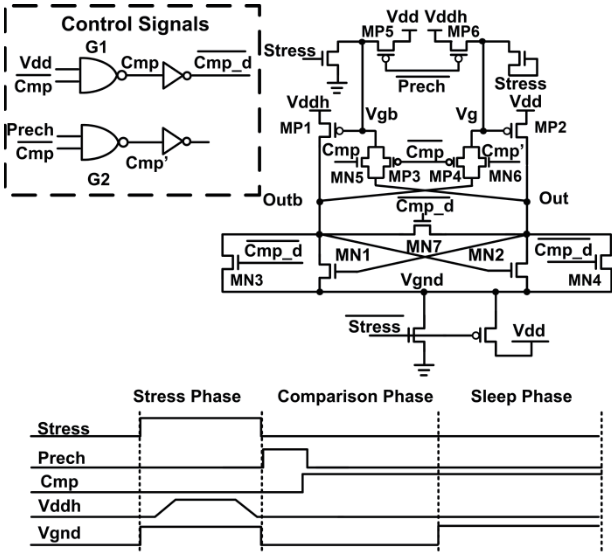


Fig. 3 Latch based aging accelerator cell.

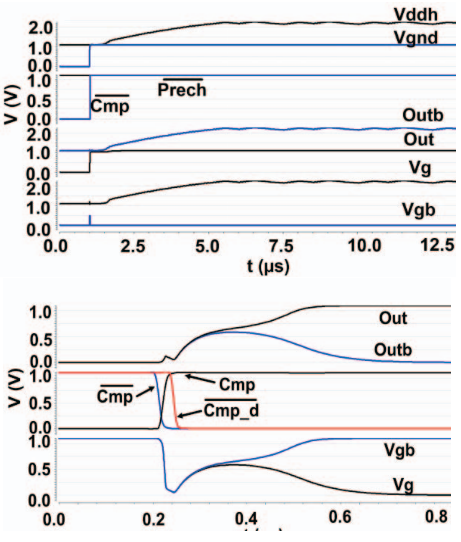


Fig. 4 Simulated waveforms of internal nodes of the aging accelerator cell. Stress Phase (upper); Comparison Phase (lower).

Table 1 Sensitivity to mismatch for transistors used in the latch.

Transistor	MP1 &MP2	MN1 &MN2	MN3 &MN4	MN5&MN6 &MP3&MP4	G1 &G2
Sensitivity	1	0.11	0.1	0.03	<0.01

After stress phase, the voltage Vddh is returned back to Vdd to perform a threshold voltage comparison between MP1 and MP2. Before comparison, the Vgnd is first released to ground. Then the out and outb are precharged to ground by MN3 and MN4 while the MP1 and MP2 are turned off from MP5 and MP6 to avoid short-circuit currents. Upon finishing the precharge of MP1 and MP2, the transmission gates (MN5/MN6 and MP3/MP4) are turned on and the transistors MN3 and MN4 are turned off to release the latch output Out/Outb from ground. The speed of rising of the nodes Out and Outb depends on the strength of MP1 and MP2 and thus establish a comparison of transistor threshold voltages. During the comparison phase, the mismatch from three supporting pairs of transistors also impact the output of the latches, i.e. MN1/MN2, MN3/MN4, transmission gates MP3/MN5 and MP4/MN6. To minimize the impact of MN3 and MN4, a shorting transistor MN7 controlled by the release signal of MN3 and MN4 is placed between Out and Outb to remove the influence from the MN3 and MN4 at the time the comparison is issued. Because the outputs are precharged to ground at the release of the latch, the impact from MN1 and MN2 is much smaller than MP1 and MP2. Besides, longer channel length (100nm) were used for MN1 and MN2 to reduce their impact on the comparison. The size of MP1 and MP2 is chosen to be 1μm/50nm to dominate the comparison. The RC delay from the transmission gates are much smaller than the comparison speed of the latch and thus their device mismatch impact from the transmission gates is also minimal. To fully characterize the mismatch impacts from the transistors, internal mismatch was generated into each transistor to perform a simulation. The mismatch impact of each transistor is compared with the target transistors MP1 and MP2 as shown in Table 1. For example, to overwrite 1mV mismatch from MP1 and MP2, 9mV mismatch (normalized to the same size) needs to be present in MN1 and MN2. The mismatch impact from all supporting transistors can be lumped into an external random noise term to quantify the disturbance it generates to the detection accuracy. Details will be described in section V. Note that all the control signals are common to MP1 and MP2 except Cmp and Cmp' whose impact is also characterized.

IV. CHARGE PUMP BASED VOLTAGE GENERATOR AND WHOLE SYSTEM

An internal generated stress voltage is needed because the supply voltage may observe variation and the users may intentionally drop the supply voltage. Therefore, a simple voltage doubler will not work for the target application. A charge pump with feedback control is proposed in this work to establish a well-controlled stress condition for the device. Fig. 5(a) shows the proposed voltage generator based on Dickson charge pump with feedback loop to monitor the output voltage. Core transistor is used and no voltage stress over Vdd is generated onto the transistors. Because during high stress period only leakage currents (subthreshold leakage and gate leakage) flow through Vddh, the charge pump only needs to support a small current flow, e.g. around 1μA in the worst case simulation. As a result, the number of stages and capacitance values (~0.5pF) are minimized to reduce the area cost. Fig. 5(b) shows the simulated voltage waveforms during the operation of the charge pump. Note that to reduce the current drawn to the Vddh from the resistor divider which provides feedback voltage to the voltage comparator, the activation of the comparator and resistor ladder is activated with only 3% duty cycle with a 2MHz monitoring clock. A low duty cycle and low clock frequency helps reduce the current that the charge pump needs to support leading to reduced size of the voltage generator. Due to the low duty cycle, a voltage ripple is observed in the stress voltage. In this case, a voltage ripple of ±50mV is observed which does not create significant variation on the stress time because the average output voltage is maintained the same.

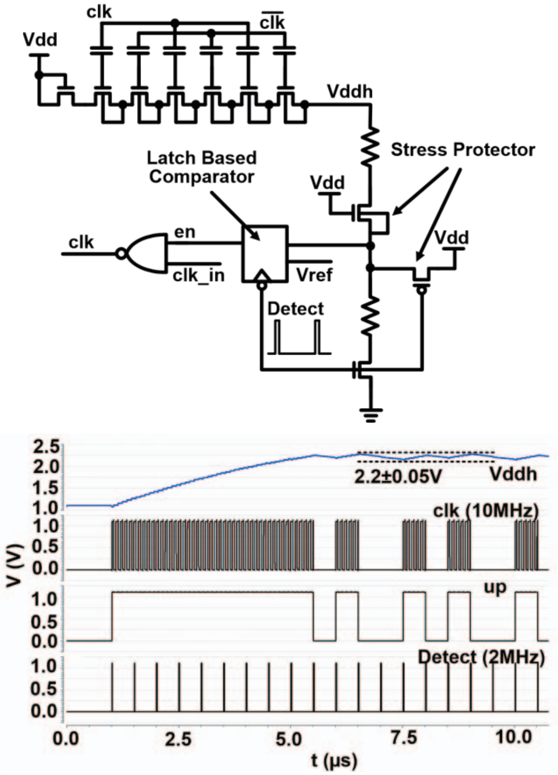


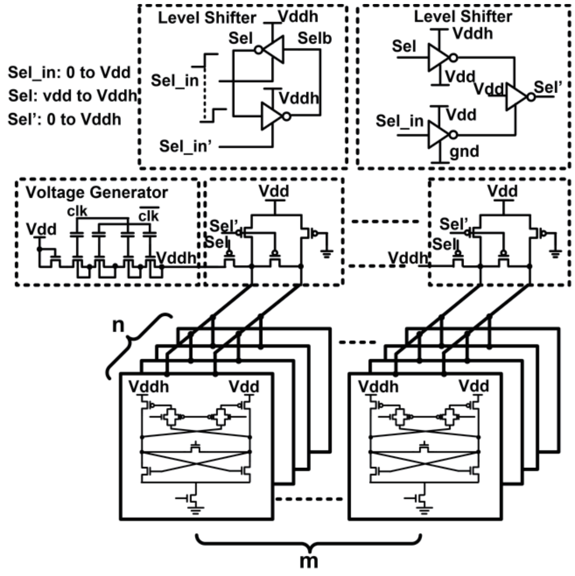
Fig. 5 Charge pump based voltage generator with feedback control (upper); Simulated operating waveform (lower).

Fig. 6(a) shows the overall system level diagram of the silicon tracking circuits. A balanced power switch is used to connect Vddh to internal Vdd during comparison phase. The comparison operations of multiple latch cells are staggered to eliminate the interference through power supply among the cells during comparison. Simulation has verified that the interaction among cells as well as the mismatch of the power switch has negligible impact to the output of the latch cells. Fig. 6(b) shows the simulated waveform during a complete operation cycle.

V. STOCHASTIC PROCESSING FOR VARIATION TOLERANCE

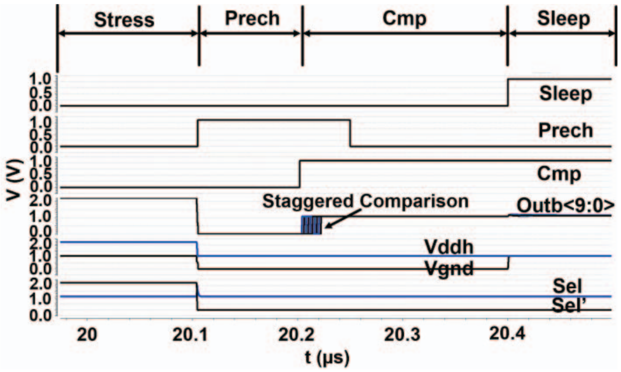
A. Stochastic Processing Methodology

Even though a noticeable threshold voltage shift has been created through the accelerator circuit described in section III, two major challenges still exist. First of all, the use of a single binary bit output from a latch cell is insufficient to distinguish a fresh IC from a used IC as the binary output from a fresh IC is random and thus cannot provide sufficient information to determine whether the chip has been used. Secondly, the enhanced threshold voltage shift introduced from the elevated stress shown is still not sufficient to guarantee a high yield. For example, with a threshold voltage shift of 21mV, still 1 out of 20 chips will fail the detection. Combining the two issues above, a stochastic based processing methodology is introduced in this work to improve the reliability of the detection. As a single cell is not able to provide sufficient information on the chip usage, the stochastic output of multiple latch based cells is use.



(a) System top level schematic

Fig. 7 explains the flow diagram of the stochastic processing methodology. An array of *m* x *n* cells are deployed on the chip. A set of *n* cells are stressed simultaneously each time the chip goes through a reboot from a Power-on-Reset (POR) circuit. Totally m times reboot can be tracked depending on the need of the application. After each reboot, a fresh row of n cells are selected to be stressed.



(b) Simulated waveforms

Fig. 6(a) the top level diagram of the silicon tracking circuits. (b) Overall operation waveforms.

After stress, the n cells will ideally display outputs of all “1” compared with an unstressed situation where the number of “1”s is random. By counting the number of “1”s at the output, the usage of the chip can be detected. However, the on-chip mismatch introduces substantial “noise” into the comparison. As a result, a stressed chip may not observe all “1” at the n bits output if any of the cells observes more mismatch than the induced threshold voltage shift. To deal with the significant mismatches, the detection rule can be relaxed to allow *t* numbers of faulty output “0” and still consider the chip has been stressed.

The relaxation leads to an optimization problem from two failure scenarios: (1). An unstressed chip happens to exhibit an output pattern with less than or equal to *t* of "0"s. Such a chip will be mistakenly considered to be stressed. We call this chip as a "missed" chip. (2). A stressed chip with more than *t* cells with larger mismatch than the introduced Vth shift. Such a chip cannot be stressed into the final pattern, i.e. less than *t* numbers of "0"s. We call this chip as a "defective" chip. The sum of scenarios (1) and (2) gives the total failure rate of the detection. The rates from "missed" chip and "defective" chip can be formulated as bellow.

The miss rate for a tolerance of *s* incorrect bits for a total of n bits is:

$$MR(s) = \frac{C(n,s)}{2^n} = \frac{n\cdot(n-1)\cdot(n-2)\ldots(n-s)}{s!\cdot 2^n} \quad (2)$$

where $C(n,s)$ is the combination formula.

Total miss rate for a maximum *t* incorrect bits is given by:

$$MR_{all} = \sum_{s=0}^{t} MR(s) = \sum_{s=0}^{t} C(n,s)/2^n \quad (3)$$

The standard deviation of mismatch between two PMOS transistors is given by σ_0. If the rest of circuits introduces an additional effective mismatch given by σ_1 as evaluated in section III, the total equivalent mismatch can be found by $\sigma_2^2 = \sigma_0^2 + \sigma_1^2$. From σ_2, we can find the possibility (denoted by *p*) of a large mismatch of all the transistors exceeding the NBTI threshold shift based on the cumulative distribution function (CDF) of a Gaussian distribution. Thus the defective rate for s bits of faulty outputs is:

$$DR(s) = p^s \cdot (1-p)^{n-s} C(n,s)$$
$$= p^s \cdot (1-p)^{n-s} \cdot \frac{n\cdot(n-1)\cdot(n-2)\ldots(n-s)}{s!\cdot 2^n} \quad (4)$$

Total defective rate for a tolerance of *t* incorrect bit is given by:

$$DR_{all} = \sum_{s=t+1}^{\infty} DR(s) = \sum_{s=t+1}^{\infty} p^s (1-p)^{n-s} C(n,s) \quad (5)$$

Note that that the summation starts from t+1 because we consider the faulty number of bits equal to *t* or less to be a successful write. Thus, the total failure rate for *n* bits with a *t* bits tolerance is given below:

$$FR = MR_{all} + DR_{all}$$
$$= \sum_{s=0}^{t} C(n,s)/2^n + \sum_{s=t+1}^{\infty} p^s (1-p)^{n-s} C(n,s) \quad (6)$$

For a chosen number of *n* cells, the failure contribution from missed cells and defective cells show an opposite trend leading to an optimum solution for the value of *t*. Fig. 8 shows the miss rate, defective rate, and total failure rate versus the allowable number of faulty output *t* with n=25 based on Monte-Carlo simulation of 10,000 chips. It is seen that an optimum value of *t* exists to provide the minimum failure rate. Fig. 9 shows the total failure rate versus number of cells *n*. As the number of *n* increases, the failure rate from the "missed" chip drops, which shifts the optimum value of *t* higher and leads to a lower total failure rate. If a certain target yield is set, e.g. three sigma yield of 99.7%, the choices of *n* and *t* can be obtain from simulation as shown in Fig. 9. In this case, simulation shows an *n* of 22 and a *t* of 4 provide an optimum total failure rate of 0.3% satisfying the yield target.

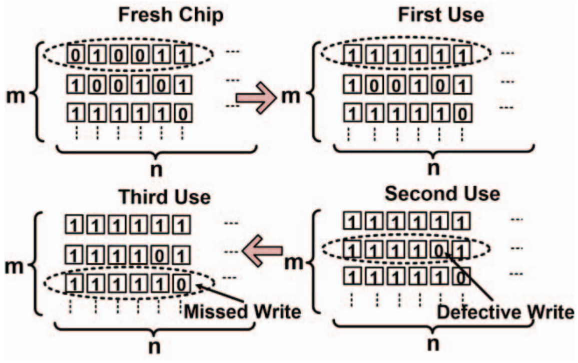


Fig. 7 stochastic processing flow for variation tolerance.

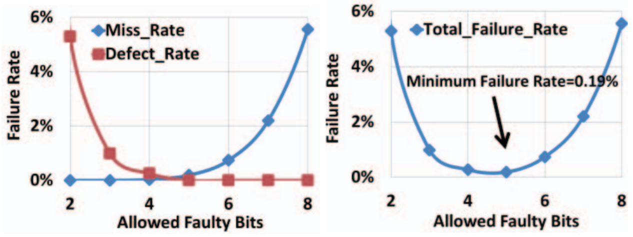


Fig. 8 the simulated failure rate versus the allowed number of "0" at the output for a fixed number of 25 cells.

Fig. 10 shows the analytical results from equations above on the minimum numbers of cells to achieve a 3-sigma yield versus different stress time. It is shown that 22 cells are needed for 2 second stress while only 9 cells are needed for 200s. Note that further increasing stress time will not reduce cell counts any longer because the failure rate is limited by the miss rate from a fresh chip, which gives a fundamental stochastic requirement independent of choices of device size, cell count, etc. Fig. 11 shows the full-schematic transistor level simulation for 5000 chips using Monte-Carlo simulation. 23 cells are used to achieve a yield of 0.3% for 2 seconds of stress time. Compared with analytical result, one more cell is needed most likely due to mismatch impact of unaccounted transistors in the schematic.

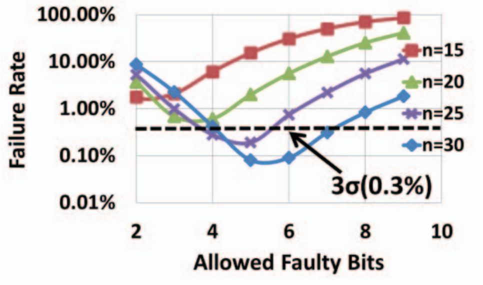


Fig. 9 the simulated total failure rate versus number of cells *n*.

The area of the latch cell is estimated to be 11μm^2, which is about 15% larger than a standard cell flip-flop. The overall area of the silicon tracking circuit is equally dominated by the latch cells and charge pump capacitors which are estimated to be about 25 latch cells. If multiple cell arrays are used, the size the charge pump generator becomes less significant as only one charge pump is needed for the whole chip. Thus, the size of the silicon tracking circuit is determined by the numbers of latch cells needed based on the yield requirement. The overall power consumption is simulated to be about 25μW during operation. It is worth to mention that the design is highly scalable with technology and the hardware cost is expected to reduce further in more advanced technology where the aging effect becomes more pronounced.

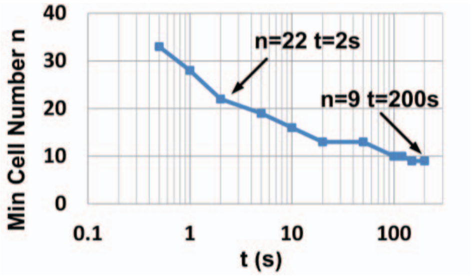


Fig. 10 Minimum cell numbers versus stress time for a 3σ yield.

VI. CONCLUSIONS

As hardware security and online configurability is becoming a critical requirement in a modern IC design, an accurate and cost efficient silicon tracking circuit is strongly desirable. This paper proposes a novel silicon usage tracking circuit using only conventional CMOS devices. By applying both device level and system level acceleration technique, the aging effect based on NBTI is enhanced to create clearly distinguishable tracker of the IC usage with tolerance of process variation. A stochastic based processing and design technique is proposed to improve the successful rate of tracking. The entire circuits were designed in a 45nm CMOS technology with Monte-Carlo simulation to verify the functionality and effectiveness of the proposed scheme. Results show that it is possible to achieve a high confident tracking within a few seconds in contrast with previous reported days of operation. Simulation also shows that the system only consumes 25μW power. The design is self-contained and the size of building elements of the proposed silicon tracking circuits is compatible with conventional standard cell leading to an easy implementation as an embedded IP.

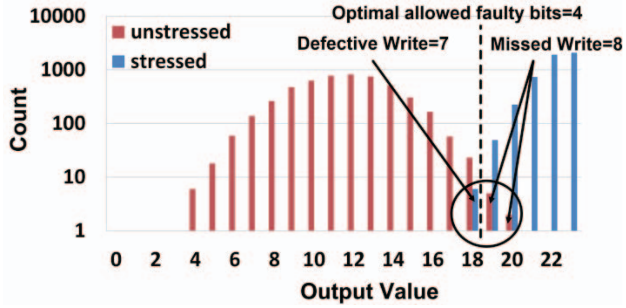


Fig. 11 Full-schematic simulation with 5000 chips verifying the analysis results with *n*=23 and stress time of 2 seconds. Total failure rate of 0.3% is observed.

ACKNOWMENTS

This work was funded by NSF grant CCF-1533656.

REFERENCES

[1] G. Contreras, T. Rahman, and M. Tehranipoor, "Secure Split-Test forPreventing IC Piracy by Untrusted Foundry and Assembly", *International Symposium on Fault and Defect Tolerance in VLSI Systems*, 2013.

[2] K. Huang, J. Carulli, and Y. Makris. "Parametric counterfeit IC detection via Support Vector Machines", *International Symposium on Fault and Defect Tolerance in VLSI Systems*, Oct. 2012.

[3] J. Stradley and D. Karraker, "The electronic part supply chain and risks of counterfeit parts in defense applications," *Trans. on Component Package Technology*, vol. 29, no. 3, pp. 703–705, 2006.

[4] L. W. Kessler and T. Sharpe. (2010). "Faked Parts Detection", http://publish-it-online.com/display_article.php?id=411055

[5] M. Hosomi, H. Yamagishi, et al., "A Novel Nonvolatile Memory with Spin Torque Transfer Magnetization Switching: Spin-RAM," *in IEEE International Conference on Electron Devices Meeting Technical Digest*, Dec. 2005.

[6] A. Raychowdhury, D. Somasekhar, T. Karnik, and V. De, "Design Space and Scalability Exploration of 1T-1STT MTJ Memory Arrays in the Presence of Variability and Disturbances," *in IEEE International Conference on Electron Devices Meeting*, Dec. 2009.

[7] T. Kim, R. Persaud, and C. H. Kim, "Silicon odometer: An on-chip reliability monitor for measuring frequency degradation of digital circuits," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 4, pp. 874–880, Apr. 2008.

[8] J. Keane, X. Wang, D. Persaud, and C. H. Kim, "An all-in-one silicon odometer for separately monitoring HCI, BTI, and TDDB," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 4, pp. 817–829, Apr. 2010.

[9] Ujjwal Guin, Xuehui Zhang, Domenic Forte, and Mohammad Tehranipoor, "Low-cost On-Chip Structures for Combating Die and IC Recycling", *Design Automation Conference*, June 2014.

[10] Xuehui Zhang, Mohammad Tehranipoor, "Design of On-Chip Lightweight Sensors for Effective Detection of Recycled ICs", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 5, pp.1016-1029, May 2014.

[11] Matthieu Deloge, Bruno Allard, et al., "Application of a TDDB Model to the Optimization of the Programming Voltage and Dimensions of Antifuse Bitcells", *IEEE Electron Device Letters*, vol. 32, no. 8, Aug. 2011.

[12] Fabrizio Torricelli, Luca Milani, Luigi Colalongo, Anna Richelli, Zsolt Miklos Kovacs-Vajna, "Half-MOS Based Single-Poly EEPROM Cell With Program and Erase Bit Granularity," *IEEE Electron Device Letters*, vol. 34, no. 12, pp. 1041-1043, Dec. 2013.

[13] Koichi Fukuda, Yoshihisa Watanabe, et al., "A 151-mm2 65-Gb 2 Bit/Cell NAND Flash Memory in 24-nm CMOS Technology", *IEEE Journal of Solid-State Circuits*, vol. 47, no. 1, pp. 75-84, 2012.

[14] Sanu K. Mathew, Sudhir K. Satpathy, et al., "A 0.19pJ/b PVT-Variation-Tolerant Hybrid Physically Unclonable Function Circuit for 100% Stable Secure Key Generation in 22nm CMOS", *IEEE International Solid-State Circuits Conference*, Feb. 2014.

[15] Mudit Bhargava, Ken Mai, "An Efficient Reliable PUF-Based Cryptographic Key Generator in 65nm CMOS", *Design, Automation and Test in Europe*, Apr. 2014.

[16] Mudit Bhargava, Kaship Sheikh, Ken Mai, "Robust True Random Number Generator Using Hot-Carrier Injection Balanced Metastable Sense Amplifiers", International Symposium on Hardware-Oriented Security and Trust (HOST), 2015.

[17] K. Miyaji, T. Suzuki, S. Miyano, and K. Takeuchi, "A 6T SRAM with a carrier-injection scheme to pinpoint and repair fails that achieves 57% faster read and 31% lower read energy," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC),* 2012.

[18] Andreas Kerber, "Metal Gate / High-k Reliability Characterization: From Research to Development and Manufacturing", *International Symposium on Advanced Gate Stack Technology*, Sept. 2010.

[19] Andreas Kerber, Siddarth A. Krishnan, Eduard Albert Cartier, "Voltage Ramp Stress for Bias Temperature Instability Testing of Metal-Gate/High-k Stacks", *IEEE Electron Device Letters*, vol. 30, no. 12, pp. 1347-1349, Dec. 2009.

[20] Rakesh Vattikonda, Wenping Wang, Yu Cao, "Modeling and Minimization of PMOS NBTI Effect for Robust Nanometer Design", *Design Automation Conference*, June 2004.

[21] S. Mahapatra, N. Goel, et al., "A Comparative Study of Different Physics-Based NBTI Models", *IEEE Transactions on Electron Devices*, Vol. 60, No. 30, pp. 901-916, March 2013

[22] Seyab Khan, Said Hamdioui, "Temperature Dependence of NBTI Induced Delay", *IEEE International On-Line Testing Symposium*, July 2010.

[23] Hong Luo, Yu Wang, Ku He, Rong Luo, Huazhong Yang, Yuan Xie, "Modeling of PMOS NBTI Effect Considering Temperature Variation", *International Society for Quality Electronic Design*, Mar. 2007.