OxID: On-Chip One-Time Random ID Generation using Oxide Breakdown

Nurrachman Liu, Scott Hanson, Dennis Sylvester, David Blaauw
University of Michigan, Ann Arbor, MI

Abstract

A new chip ID generation method is presented that leverages the random and permanent characteristics of oxide breakdown. A 128b ID array is implemented in 65nm CMOS and two algorithms for stressing the oxides are presented, showing a near-ideal Hamming distance of 63.92 in silicon measurements and consistent IDs across voltage and temperature.

Introduction

Chip ID systems are used to enforce user licenses as well as in communication and security protocols. In these applications, it is desirable to generate IDs on-chip at the application point so IDs are guaranteed unknown until first used. This avoids the need for off-chip, pre-generated IDs that are programmed using fuses, a process that exposes IDs to human intervention and storage on computers that may be compromised.

A key requirement for chip ID generation is that the generated ID is unique to only that chip, and that the ID is time and environmentally invariant. The chances that two chip IDs have all, or many, bits the same is minimized by using a large bit width (e.g., 128bits/ID) and ensuring a high degree of randomness during generation. Previous methods rely on inherent threshold voltage (V_t) mismatch between devices, which is detected by measuring either device current [1] or inherent SRAM bitcell skew towards 0 or 1 states [2]. However, V_t mismatch can be very small between any particular transistor pair, making it difficult to repeatedly generate an identical ID for a given chip. Hence, previous approaches exhibit a small number of bit flips between successive ID readings (i.e., the IDs had a non-zero self Hamming distance), complicating the use and reliability of chip IDs.

This paper presents a new method called OxID that generates chip IDs using oxide breakdown. We leverage the fact that oxide breakdown is an inherently random effect [3] (one oxide may break long before another identical oxide under the same stress conditions) and is also both abrupt and permanent. Hence, it enables improved ID stability over time and environmental conditions. Once an oxide breaks down, its resistance changes from a nearly infinite value to the order of MΩ or kΩ [4], which has made it popular for one-time-programmable arrays [5,6]. Silicon measurements of 162 ID generators in this work demonstrate nearly ideal randomness of the generated IDs, maximizing their uniqueness. The proposed approach can also detect prior ID generation; if on first use the ID is non-zero, this indicates that the ID was previously generated through possible intrusion and may be compromised.

Proposed System and ID Generation Method

OxID consists of a memory array composed of 3-T memory cells that use a thin-oxide moscap as a fuse element (Figs. 1 and 2). The array has 16 rows by 8 columns, totaling 128 cells, each of which can be read through a bitline and sense amplifier. All oxides in the array are exposed to a stress voltage of 4.5V and identical stress time. While the breakdown times of the 128 oxides follow a random distribution, the mean of this distribution is difficult to determine *a priori* and can change from chip to chip due to oxide thickness variations. Hence, exposing all chips to a preset stress time will likely result in a significant portion of OxIDs with oxides either all broken or unbroken. Therefore, we propose two algorithms that dynamically adjust stress time to ensure that close to half of all oxides break while half remain intact. Both algorithms stress the array in small time increments using an on-chip controller. In the first algorithm, the entire array is read out after each stress interval. Initially, the array will read nearly all zeros and gradually contain more ones as oxides start to fail. When ones exceed zeros, the stress iterations are terminated and the ID is complete. By dynamically checking the array state after each stress interval, the algorithm automatically adapts to the global condition of the oxides, providing added stress to more inherently reliable arrays. It also provides immunity to voltage fluctuations during the stressing.

One drawback of this approach is that all generated IDs will have a nearly identical number of zeros and ones. While this does not reduce the randomness of the IDs, it does reduce the number of possible ID permutations. For a 128b ID, the number of possible ID permutations reduces by a factor of $\sim 2^{3.8}$. Hence, if an equivalent pool of IDs is required as in a standard random ID, the number of bits must be increased (e.g., for a 128b ID, by 4 bits or ~3%). Therefore, we propose a second algorithm that uses a small set of canary cells to predict the number of stress iterations for the entire array. In this case, only cells specified as canary cells are read out after each stress interval and further array stress is terminated when 50% of the canary cells are broken. Due to random variation between the canary cells and the remainder of the array, a larger set of ID permutations is generated.

Both algorithms are implemented and compared. After the ID is generated using either algorithm, a final "afterburn" phase is performed where all broken oxides are strongly stressed for a longer duration. Due to limitations of stress isolation a few borderline oxides may break down as well. Hence, this process sacrifices a small Hamming distance degradation (measured at 2-3%) for higher read operation robustness across environmental conditions.

The 3-T bitcell (Fig. 3) consists of a thin-oxide SVT transistor driven by the wordline, a thick-oxide 2.5V I/O "blocking" transistor, and a thin-oxide SVT transistor with S/D tied as a moscap. This bitcell is similar to the 3-T cell in [4,7]. The thick-oxide transistor separates the thin-oxide wordline device from the high voltage of the moscap during stress. For unbroken oxides there is by design a small voltage that accumulates across the moscap due to its high leakage at high V_{DD} (0.7V for V_{DDH} of 4.5V). This protects oxides that have not been selected for stress. Cell currents are limited by the resistance of the minimum-sized thick-oxide transistor and word-access transistor. During cell read, V_{DDH} is shorted to V_{DD}. For experimentation, the 128 oxides can be stressed all at once or by row, column, or cell.

Measurement Results

OxID was implemented in a standard 65nm CMOS technology. For experimentation, the gate voltage for the blocking transistor (VBT) and the sense amplifier reference voltage were brought in from off-chip, but can also be generated on-chip. We applied the global stress algorithm described above at room temperature to 162 arrays and the canary-based algorithm to 144 arrays. Two perfectly random IDs should, on average, have a Hamming distance of exactly half the total number of bits in the ID. Comparing all pairs of ID bit sequences (13041 and 10296 pairs, respectively), the average Hamming distance for the global algorithm is 63.92, close to the ideal value of 64 (Fig 4). The average Hamming distance for the canary algorithm is 61.79, implying a trade-off in randomness and ID set size (Fig 5). The read power is 0.34 pJ per bit (Table 1). The self-Hamming distance upon repeated reading of the ID in different environmental conditions was tested for 14 arrays. Results show 0 self-Hamming distance for up to 100mV supply voltage deviation from 1.1V nominal and across temperature from 0°C to 85°C. Figs. 7 and 8 show the self-Hamming distance as a function of voltage and sense amplifier read margin across temperature. Fig. 9 shows the generated bits for each cell location, averaged across all arrays, with no obvious spatial artifacts. The spatial distribution of the breakdown time of each oxide in a typical array is shown in Fig 6. Table 1 provides a comparison of OxID to related prior work [1,2], showing improved energy, stability, and density. Fig. 10 shows the number of stress intervals across all arrays. Fig. 11 shows the chip microphotograph and chip statistics are included in Table 2.

Acknowledgements

We thank STMicroelectronics for test chip fabrication support.

References

[1] K. Lofstrom, *et al.*, *ISSCC*, pp. 372-373, 2000.
[2] Y.Su, J.Holleman, B. Otis, *ISSCC*, pp. 406-407, 2007.
[3] J. Stathis, *J. of Applied Physics*, pp. 5757-5766, Vol. 86, Nov. 1999.
[4] J. Kim, and K. Lee, *Electron Device Letters*, pp. 589-591, Sept. 2003.
[5] P. Candelier *et al.*, *IRPS*, pp. 169-173, 2000.
[6] H. Ito, and T. Namekawa, *CICC*, pp. 469-472, 2004.
[7] H-K Cha *et al.*, *JSSC*, pp. 2115-2124, Vol. 41, No. 9, Sept. 2006.

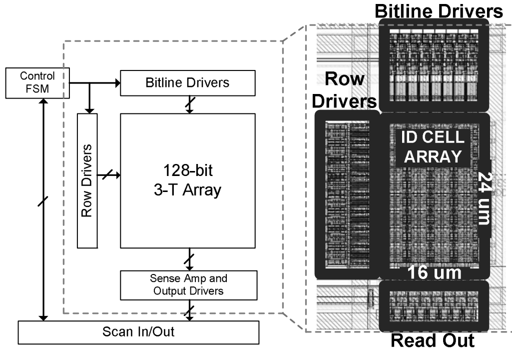


Figure 1: System architecture.

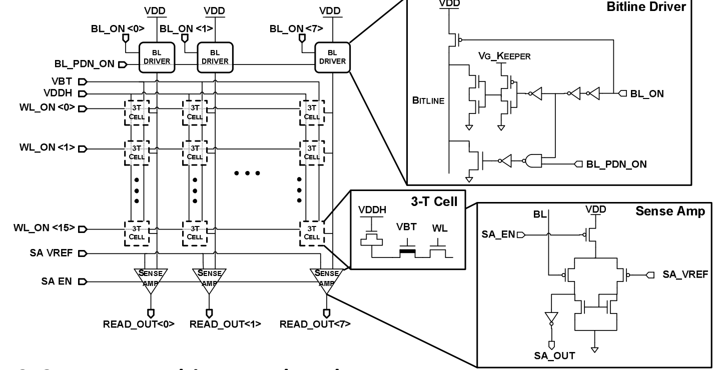


Figure 2: 3-T ID Array, drivers, and readout.

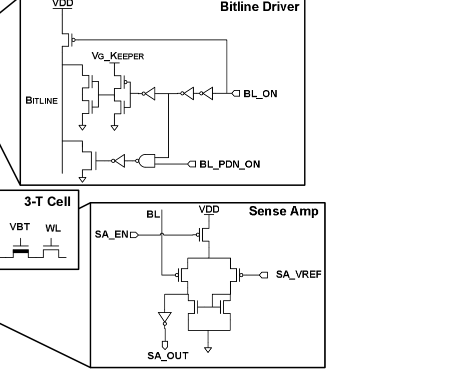


Figure 3: 3-T cell; bitline driver; sense amp

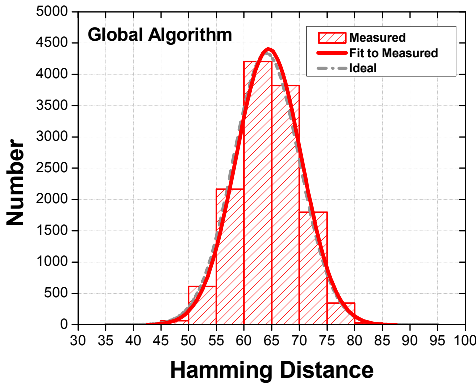


Figure 4: Measured Hamming distance for global algorithm. 162 Arrays total.

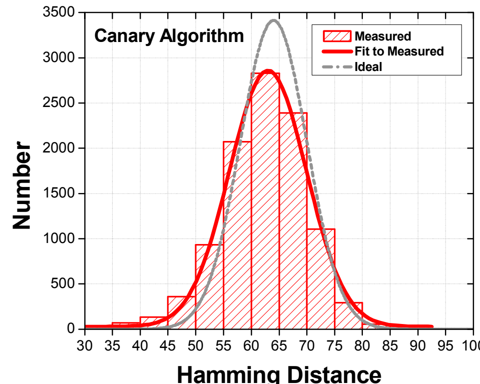


Figure 5: Measured Hamming distance for canary algorithm. 144 Arrays total.

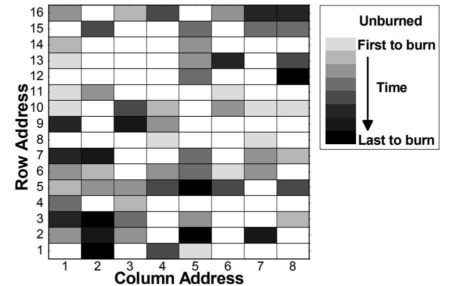


Figure 6: Measured spatial time evolution of a typical array oxide burn times.

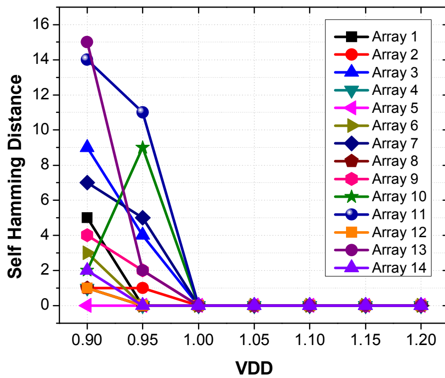


Figure 7: Measured # of changing bits over power supply sweep.

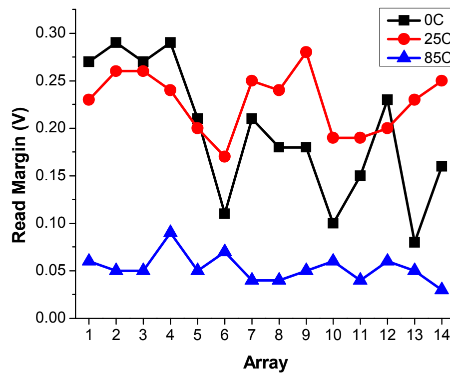
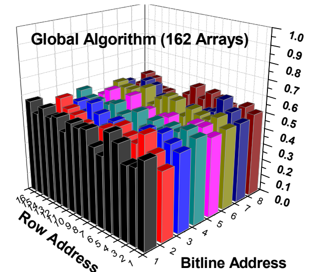


Figure 8: Measured read margins across temperature.



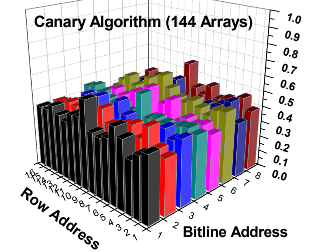
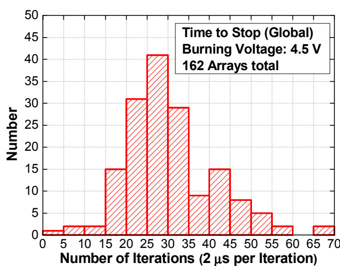


Figure 9: Spatial dependency: Global alg. (top); Canary alg. (bottom)



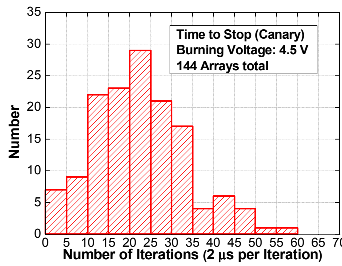


Figure 10: Distribution of stress intervals for all arrays. Global algorithm (left); Canary algorithm (right)

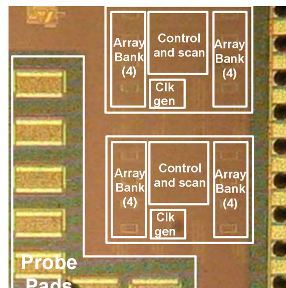


Figure 11: Die photo (65nm).

	Global Algorithm
VDD	1.1 V
Bit Length	128
Throughput (Bps)	625 M
Average Hamming Distance (162 arrays)	63.92

Table 2: Chip statistics

Table 1: Comparison with previous work

	Through-put (Bps)	Energy per bit (pJ/bit)	Average Unstable Bit	ID Length	Tech-nology (nm)	Area (um^2)	Area Scaled to 65nm (um^2)
[1]	3.75k	8330	N/A	112	350	23,496	939.84
[2]	125k	0.93	3.9	128	130	15,288	3822
This work	625M	0.34	0	128	65	1242	1242