A True Random Number Generator using Time-Dependent Dielectric Breakdown

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

Abstract

A true random number generator (tRNG) is proposed that, for the first time, uses the random physical process of time to oxide breakdown under voltage stress. Time to breakdown is repeatedly measured with a counter and serialized into a bitstream. The 1200 μm^2 tRNG, called OxiGen, was fabricated in 65 nm CMOS, passes all 15 NIST randomness tests without post-processing and in a 3 month run generated sufficient bits for worst-case expected internet use while being < 10% exhausted.

Introduction

Random number generation plays a crucial role in cryptography and security. For example, public key cryptography systems demand strong key pair generation to ensure a third-party cannot decrypt secret messages. Traditionally random bit sequences are generated in digital systems using pseudo-random number generators, which produce sequences that are not truly random but contain exploitable patterns, such as repetition and correlation. True random number generators (tRNGs) use physical phenomena as a randomness source. Previous on-chip tRNG architectures have used telegraph noise [1] and thermal noise as the physical source [2-8]. Thermal noise is often used indirectly with a metastable inverter [2-5], a jitter-prone oscillator [6,7], or a discrete-time chaotic pipelined structure [8]. Another approach used fluctuating gate oxide current after soft breakdown (SBD) as a noise source [9]. However, most prior architectures [1,2,4,7,8] rely on post-processing to remove bias in the generated stream, such as a "Von Neumann Corrector", which brings into doubt the randomness of the initial bit steam. Architectures that do not require a post-processor [3,5,6,9] have only passed 7 of 15 statistical randomness test in the NIST 800-22 benchmark [10], the accepted standard test for true randomness.

OxiGen Operation

We present a novel tRNG architecture based on a random physical phenomenon not previously used for on-chip random number generation. The architecture, called OxiGen, repeatedly forces soft gate oxide dialectric breakdown under voltage stress and uses the time to breakdown to generate a random bit sequence. This time to failure (TTF) is on the order of milliseconds, which is large compared to circuit speed. Hence, a key advantage of the approach is that TTF can be easily and accurately captured using a simple counter. This avoids the complex methods needed to capture thermal noise (the most common source of randomness), which is on the order of μV [3] and requires sensitive amplifiers.

The OxiGen method was implemented on an array of 128 MOS capacitors and can generate an estimated >5 billion true random bits before permanent breakdown of the capacitor oxides. OxiGen is the first tRNG to pass all 15 NIST tests without a post-processor. OxiGen was fabricated in 65nm CMOS, consumes 2 mW of power and 0.0012 mm^2 of area, and produces 11 kb/s of random data.

Breakdown occurs when a stress voltage is applied across an oxide, and is an inherently random process: for two completely identical oxides under the same stress conditions, TTF is different and unpredictable [11]. The proposed architecture measures oxide TTF using a binary counter and generates a bitstream based on the resulting value (Fig. 1). Since TTF is normally distributed with finite variance, the counter value will have similar characteristics, whereas a generated random bitstream requires a uniform distribution of bits. The method resolves this by truncating the counter value, discarding *n* high order bits while keeping remaining lower order bits and outputting them serially to create a random bitstream. First, the high bit of the binary counter value is searched for, starting from the most-significant bit position. This most significant '1' bit is discarded, along with *n-1* successive higher-order bits. The remaining lower-order bits are guaranteed to have rolled over at least 2^{n-1} times. Thus the algorithm intelligently truncates bits to maximize the number of bits generated while guaranteeing a high-quality random bitstream. This also adjusts for shifts in the mean count value due to variation in oxide thickness and wear-out conditions. Unlike post-processing, the algorithm does not observe or manipulate bits placed in the random bitstream, instead it judges the quality of bits by their position in the counter.

OxiGen relies on the fact that oxide breakdown occurs in stages. By carefully monitoring the bitline voltage, millions of random bits can be generated from each device by repeated stressing until the oxide fully breaks down. The array of 16 wordlines and 8 bitlines is made up of 128 basic 3-T cells [12,13] constructed from a MOS capacitor under stress, a thick-oxide high-voltage protection transistor, and access transistor. Each cell is stressed in turn while the bitline is weakly held by a keeper, and the bitline voltage is monitored by a simple comparator. During stress, a 30-bit counter runs at 325 MHz. As the oxide resistance drops, the comparator detects a voltage rise on the bitline and stops the high-frequency counter and disables the cell wordline to discontinue oxide stress. The counter value is read-out, truncated as described, and output as the random bitstream. A stress voltage is applied to the next cell in the array and the process repeats. Total bit generation can be increased by adding more cells to the array, and bitrate can be increased by adding more comparators and counters for parallel count generation. This implementation shares one counter for 8 bitlines, while 8 counters would yield 8× higher throughput.

To maximize the total number of random bits generated by an array, OxiGen uses an algorithm to dynamically tune the stress voltage *vddh* and comparator reference voltage *sa_vref*. As cells in the array are cycled through and stressed, TTF is evaluated and *sa_vref* is adjusted accordingly. If TTF is too short, the reference voltage is increased to generate longer counts, resulting in more random bits. Conversely, it is reduced if the reference voltage is found to be too high, causing the cell to be over-stressed without generating many additional random bits, wasting useful device lifetime. When *sa_vref* reaches it maximum value, the stress voltage *vddh* is reduced and *sa_vref* is reset to its minimum value. Thus, the algorithm dynamically converges on optimal conditions to generate the largest number of random bits, ensuring maximum cell harvest.

Measured Results

Two sets of random bit sequences generated by OxiGen were analyzed by the NIST 800-22 test suite for randomness. First, 300 sequences of 43k bits (13M bits total) were analyzed with different truncation lengths for 10 of the 15 tests. Fig. 3 shows pass rates for NIST tests as a function of bits truncated, along with the minimum pass rate to be considered random and π's pass rate. When at least one bit is truncated, OxiGen easily exceeds the minimum pass rate and is comparable to π's pass rate. The remaining 5 NIST tests require a larger sequence length and were tested with 100 sequences of 1M bits each (100M bits total). OxiGen exceeded the minimum pass rate for the large sequence length tests and overall passed all 15 NIST tests with statistical significance and without post-processing.

Fig. 4 shows automated stress and reference voltage tuning behavior as observed while generating 500M bits over a 3 month period. The array was not exhausted at the end of the 3 month period. The comparator reference voltage had incremented less than half of its range and the stress voltage was still at the initial value of 3.5V across all cells, indicating the array was still in the early stages of life after generating 500M bits.

A typical internet secure session link (SSL) key is 256 bits. Assuming a new secure internet session every 5 minutes for 24hrs/day, 500M bits provides 10 years of operation, exceeding the lifetime of most mobile devices even if the bitstream is aggressively truncated during generation.

Furthermore, an accelerated test was performed on five cells, where the oxide was stressed 1000× longer than necessary, to predict the total number of bits that can be harvested from the cell array. Fig. 5 shows the algorithm self-adjusting after a low bit harvest and then reconverging on the next optimal stress/reference voltage pair for maximum bit harvesting. Based on the progression of voltage tuning settings, the test showed at most only 5-10% of the array had been exhausted after 500M generated bits, indicating the array can generate >5B bits before permanent wear-out and giving an even greater safety factor compared to typical use.

A comparison of OxiGen with prior tRNGs is given in Table 1. A die photo of the fabricated chip is shown in Fig. 6. In summary, OxiGen passes all NIST 800-22 randomness tests, which is a first for tRNGs not using a post-processor corrector.

Acknowledgement

STMicroelectronics is gratefully acknowledged for IC fabrication.

References

[1] R. Brederlow et al., ISSCC, 2006.
[2] J. Holleman et al., JSSC, May 2008.
[3] C. Tokunaga et al., IEEE Journal of Solid-State Circuits, January 2008.
[4] D. Kinniment et al., ESSCIRC, September 2002.
[5] S. Srinivasan et al., VLSI, 2010.
[6] M. Bucci et al., IEEE Trans. on Computers, April 2003.
[7] C.Petrie et al., IEEE Trans. on Circuits and Systems, May 2000.
[8] F. Pareschi et al., ESSCIRC, September 2006.
[9] S. Yasuda et al., JSSC, August 2004.
[10] National Institute of Standards and Technology, Pub 800-22, 2001.
[11] J. Stathis, J. of Applied Physics, vol. 86, pp. 5757-5766, Nov. 1999.
[12] N. Liu et al., VLSIC, 2010.
[13] J. Kim and K. Lee, IEEE EDL, pp. 589-591, September 2003.

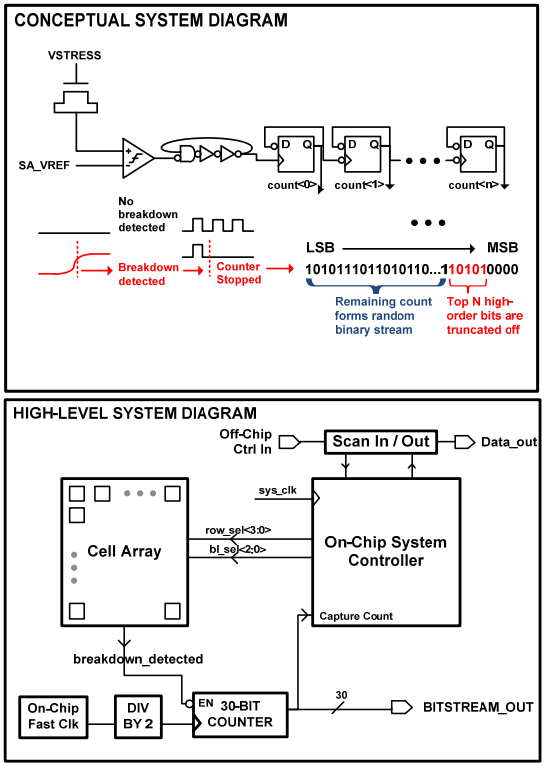

Figure 1. Conceptual system diagrams of OxiGen.

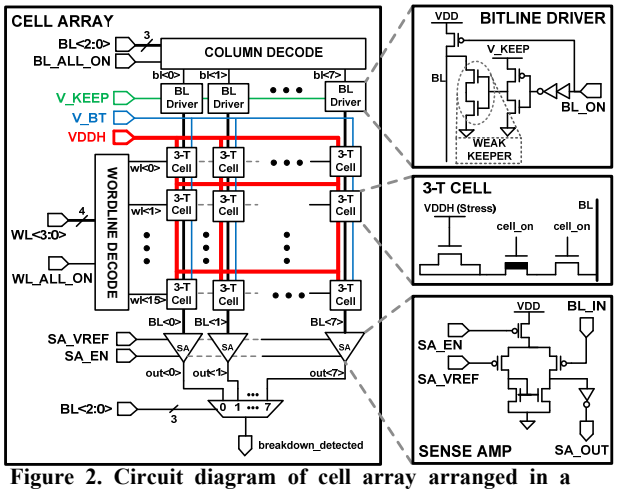

Figure 2. Circuit diagram of cell array arranged in a traditional memory configuration, showing 3-T cells, sense amplifiers, and bitline drivers.

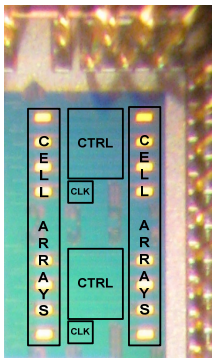

Figure 6. Die photo of OxiGen.

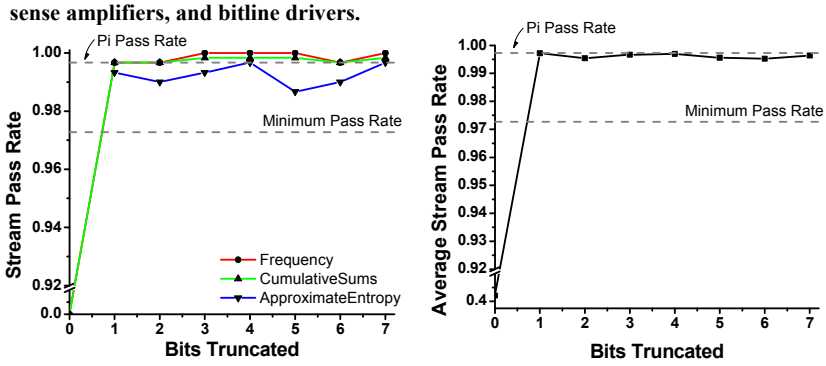

Figure 3. Stream pass rates of 300 OxiGen bit sequences for individual tests (left) and average of all 300 sequence tests (right).

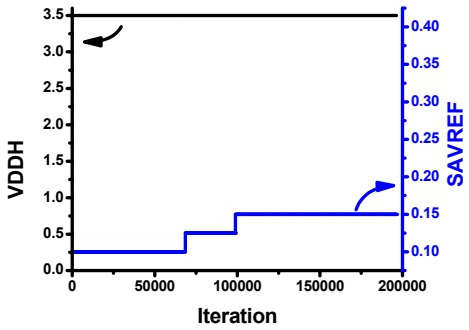

Figure 4. S*a_vref* and vddh for single cell during generation of 500Mb data. A typical iteration yields 20 bits.

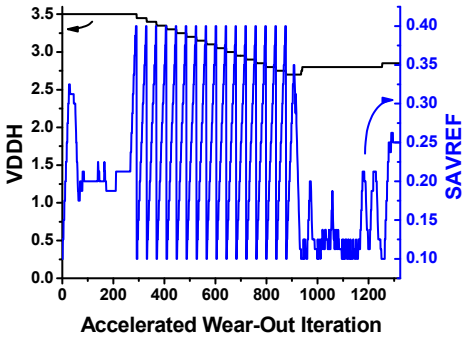

Figure 5. Accelerated stress testing diagrams.

Table 1. Table comparing OxiGen to prior true random number generators.

Ref.	Bitrate	NIST Passed	Tech.	Area (mm²)	Area Norm. (mm²)	Power (mW)	Post-Process
This work	11 kb/s	All	65 nm	0.0012	0.0012	2	No
[5]	2.4 Gb/s	7	45 nm	0.004	0.0083	7	No
[2]	5 kb/s	*	0.35 µm	0.031	0.0011	0.0094	Yes
[3]	200 kb/s	7	0.13 µm	0.145	0.0363	1	No
[1]	50 kb/s	–	0.12 µm	0.009	0.0026	0.050	Yes
[8]	40 Mb/s	All	0.35 µm	0.52	0.0179	29	Yes
[9]	50 kb/s	–	Discrete	–	–	–	No
[6]	10 Mb/s	3	0.18 µm	0.016	0.0021	2.3	No
[4]	2 Mb/s	–	0.6 µm	–	–	–	Yes
[7]	1 Mb/s	4	2 µm	1.5	0.0016	3.9	Yes

* = Though the author reported NIST results, sample sizes were insufficient for statistical significance.