WHITE PAPER

Krishnaswamy Ramkumar

V. Prabhakar

Igor Kouznetsov

Sam Geha

Cypress SONOS Technology

Abstract

This white paper discusses the key features of Cypress SONOS technology, its performance, and reliability. It also presents a comparison between the SONOS technology and other flash technologies.

Introduction

The demand for embedded flash memory has grown steeply in recent years as newer applications evolve in consumer electronics (touch screens, smart cards, bank cards, mobile payment, e-passport, etc) in addition to industrial system-on-chip (SOC) designs. The memory content of these applications has increased due to greater requirements on system performance. Despite floating-gate flash memory integration being complex with additional 9 to 12 masking steps, the industry has accepted this as the cost of having embedded nonvolatile memory. In addition, standard floating-gate flash technology has difficulty matching the baseline device models and being compatible with existing IP.

SONOS (Silicon-Oxide-Nitride-Oxide-Silicon) offers a simpler approach for embedded flash memory at a significantly lower cost. SONOS has been known as a nonvolatile memory technology since the 1980s. However, in the past it was not very successful in competing against the floating-gate technology due to its requiring a higher programming voltage. The programming voltage has been reduced in recent years by improvements in thin-film deposition technology.

Another weakness SONOS technology encountered was poor data retention at high temperatures. This problem has been solved by Cypress using charge trap engineering. Cypress's SONOS has been demonstrated to meet 10 year retention at 85°C with robust margins.

This document provides an overview of the Cypress's SONOS technology, its features, and its integration into a logic process flow.

Flash Memories

Nonvolatile memories (NVM) retain stored information even if the power supply to the memory is switched off. There are different types of nonvolatile memories including flash, read-only memory (ROM), one-time programmable (OTP), and multiple-time programmable (MTP) memories. Flash memory is the most versatile because it can be programmed and erased thousands of times with minimal degradation in the sense margin and data retention performance.

Floating Gate Flash Memory Technology

Flash memory technology has evolved over the last 30 years driven by the demand for higher density and improved performance. Figure 1 shows the cross-section structures of key flash memory cells in use today. The charge storage layer is the same in all the generations. A floating gate in a metal-oxide-semiconductor field-effect transistor (MOSFET) is used to store the charge. The floating gate is sandwiched between the channel and the control gate. It is separated from the channel by a tunnel oxide and from the control gate by an ONO dielectric in most cases. The right-most cross-section in Figure 1 (SST's cell) does not have ONO dielectric. The method of injecting (programming of memory) and removing (erasing of memory) charge from the floating gate can be different and has led to different families of flash memory.

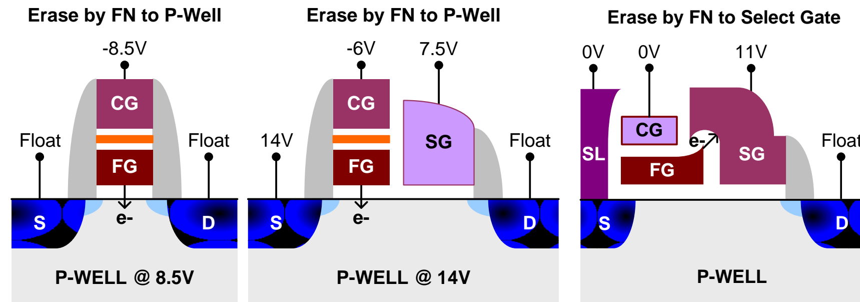

Figure 1. Cross-section Structures of Flash Memory Cells

Disadvantages of Floating Gate (FG)

The scalability of floating gate nonvolatile memory (NVM) technology is becoming challenging and expensive beyond the 55-nm node. Shrinking the device dimensions results in decreasing the supply voltage for the logic circuits. However, the program/erase voltages of the floating gate memory cannot scale because the tunnel oxide thickness cannot be thinned down without impacting data retention. This forces the use of a thick third gate oxide in the process flow to support the program/erase circuitry. This thick third gate oxide causes significant changes to core device characteristics leading to difficulties in maintaining IP compatibility with the logic/mixed-signal baseline.

Another issue is the high cost of integrating floating-gate flash memory because it needs up to 9-12 additional masks. In addition, due to the deposition of multiple poly layers needed for embedded floating-gate technologies, the defect density in the embedded flash technology becomes substantially higher than that of the baseline.

Integration of floating-gate technology with advanced High-K metal gate logic processes poses even greater challenges.

A well-known issue with floating-gate memories is that any defect in the tunnel oxide acts as a leakage path for the entire floating gate (because it is a conducting poly layer) thereby leading to weak bits. This failure mechanism needs extensive screening at wafer testing.

SONOS (or Charge Trap) Memory Technology

In contrast to the floating-gate approach, the SONOS memory uses an insulating layer such as silicon nitride with traps as the charge storage layer. The traps in the nitride capture the carriers injected from the channel and retain the charge. This type of memory is also known as “Charge Trap Memory.” Since the charge storage layer is an insulator, this storage mechanism is inherently less sensitive to the tunnel oxide defects and is more robust for data retention. The key advantage of SONOS technology is the significantly lower voltages required for program/erase operations compared to floating-gate flash. It was previously thought that the retention of a SONOS device is poor on account of the thin tunnel oxide in the ONO stack. In Cypress’s SONOS, by controlling the deposition parameters in the ONO formation, the ONO stack has been engineered so as to maximize the charge trapping efficiency during erase and program operations and minimize the charge loss during retention. With such a stack, Cypress’s SONOS meets the AEC automotive grade 1 reliability specs.

Overview of Cypress’s SONOS Technology

Cypress’s SONOS is an embedded nonvolatile memory technology that integrates a highly reliable SONOS transistor into a CMOS process flow with minimum additional mask layers and minimal impact on the electrical parameters of the existing CMOS FETs. With simple integration compared to floating-gate flash, Cypress’s SONOS technology enables a faster manufacturing ramp to high product yield. It is compatible with CMOS logic, which makes it ideal for embedded flash. This technology is qualified in Cypress’s manufacturing plant and in multiple foundries at 0.35-µm, 130-nm and 65-nm nodes. Cypress’s SONOS has been in production since 2001 and over 2 billion chips have been shipped to customers so far. The 0.35-µm and 130-nm SONOS technologies are currently in high-volume manufacturing and 65-nm SONOS technology is fully qualified in a leading foundry. This technology is also being scaled to more advanced technology nodes such as 55 nm, 40 nm and 28 nm in multiple foundries. The SONOS cells in these technologies have been shown to meet the reliability specifications and flash macros using these cells are currently undergoing qualification.

Cypress’s SONOS Transistor and Cell

The heart of the Cypress SONOS technology is the SONOS FET shown in Figure 2. This is a MOS transistor with ONO stack as the gate dielectric. The ONO stack is designed to provide the required program or erase speeds and excellent reliability. The trap distribution in the stack is engineered to yield a memory cell with outstanding endurance and retention characteristics. Furthermore, the stack is suitable for high-volume manufacturing with high process capability. The SONOS transistor shares many of the key process steps with the CMOS transistors. Hence, many regions of the SONOS transistor, such as source, drain, and gate, are identical to those of the CMOS transistors. This makes the process architecture of the embedded SONOS technology significantly simpler than that of floating gate.

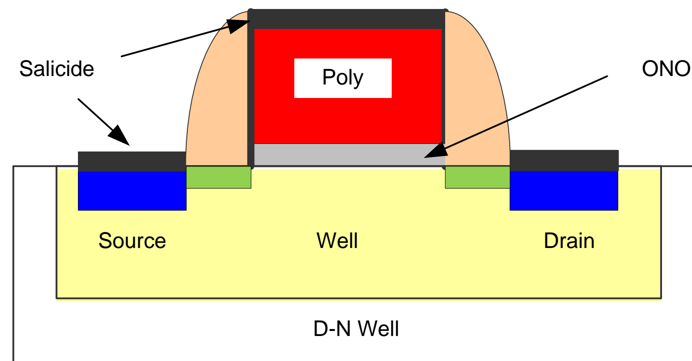

Figure 2. SONOS Transistor Cross-Section

Figure 2 is a SONOS transistor that is fabricated using a typical logic CMOS process flow. In the advanced foundry technology, the SONOS transistor has additional features such as stress nitride layer and ultra shallow junctions. The Cypress SONOS technology currently offers multiple cell options to fit into different applications, trading off cell and macro sizes for access time. Similarly on the most advanced technology nodes that use High-K dielectric and Metal Gate in the gate stacks, the SONOS transistor will exploit all the key benefits from these new features.

SONOS Cell Operation

The Cypress SONOS cell is a two-transistor (2T) cell with a SONOS transistor in series with an NMOS select transistor. The cell is programmed by raising the gate voltage of the SONOS transistor to the required positive value. Electrons are injected from the substrate into the charge-storage layer of the ONO stack by Fowler-Nordheim (FN) tunneling. The device is erased by applying the required negative voltage to the gate, which causes FN tunneling of holes from the substrate to the charge-storage layer. The required program and erase voltage between the gate and substrate is obtained by applying appropriate voltages to the gate and the p-well. Typical program and erase voltages for a 130-nm SONOS technology are shown in Table 1. Currently, Cypress uses 10 V (+7 V/-3 V) in the 130-nm technology and 7.5 V (+4 V/-3.5 V) in the 65-nm technology and more advanced nodes all the way down to 28 nm.

	CG	BL	SL	P-Well
Erase	-3.V	7.0V	7.0V	7.0V
Program	7.0V	-3.0V	-3.0V	-3.0V
Read	0V	1.2V	0V	0V

Table 1. Voltages for 130-nm SONOS Operation

A typical program and erase characteristic of the Cypress SONOS device at 85 °C is shown in Figure 3. High voltages for program and erase are generated by on-chip charge pump circuits.

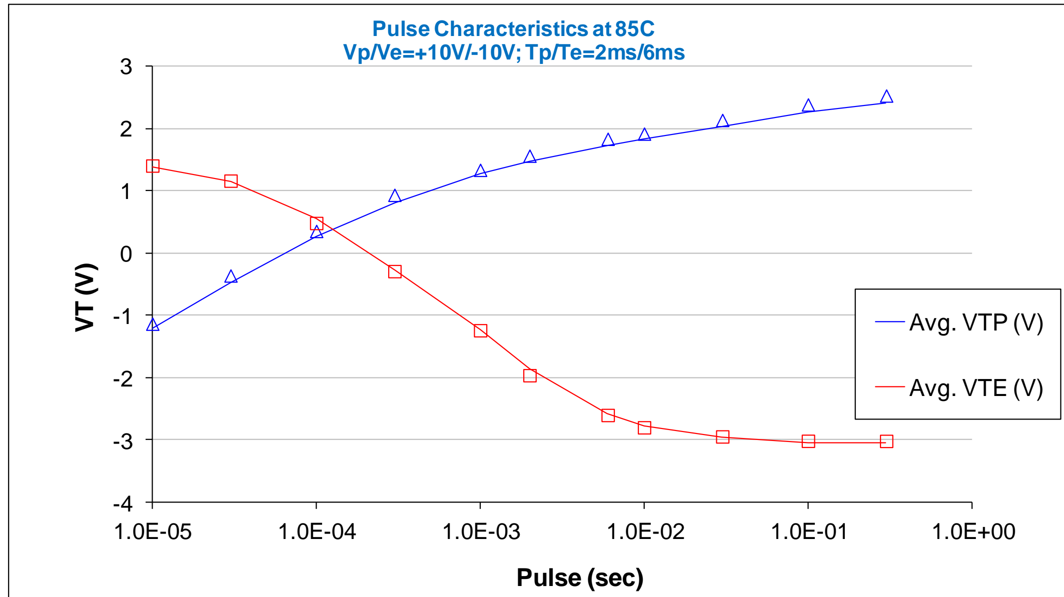

Figure 3. Program and Erase Characteristics of Cypress SONOS FET

The current program speed is 1-3 ms and erase speed is 2.5-10 ms depending on cell option and macro architecture.

In response to the customer requirement of smaller flash macro size, Cypress has now introduced a new NVM cell known as “Common Source Line (CSL) cell“. This cell has one common source line for an entire row of cells instead of the conventional “Dedicated Source Line cell”, which has one source line for each column of cells. This change makes the cell size significantly smaller and enables reduction of the flash macro area. The 55-nm and 40-nm SONOS technologies use the CSL cells for all flash macros.

Cypress’s 130-nm SONOS Performance Specifications

The optimized Cypress SONOS technology guarantees the following specifications:

Program time: 1-3 ms; Erase time: 2.5-10 ms

Program Vt (BOL): 1.5 V; Erase Vt (BOL): –1.5 V

Minimum selected erased cell read current: 8 µA/cell

Maximum selected programmed cell read current: 100 nA/cell

Maximum unselected cell leakage at 100°C: 4 nA/cell

Data retention: 10 years at 85°C

Endurance: 100k cycles

SONOS Cell Reliability

Key characteristics of an NVM cell are endurance and data retention. Endurance is determined by cycling a SONOS cell through the required number of program/erase cycles and measuring the shift of program and erase Vts. Data retention is determined by taking the SONOS cell through a fixed number of program/erase cycles and then measuring the change of Vt (program or erase) with time at an appropriate elevated temperature. The End-of-life (EOL) Vt window is determined by the degradation caused by program/erase cycles (Endurance) and Vt decay during storage (Data retention). Cypress's 130-nm SONOS technology guarantees retention of 10 years at 85°C ambient temperature after 100k cycles. Figure 4 and Figure 5 illustrate the typical endurance and data retention characteristics.

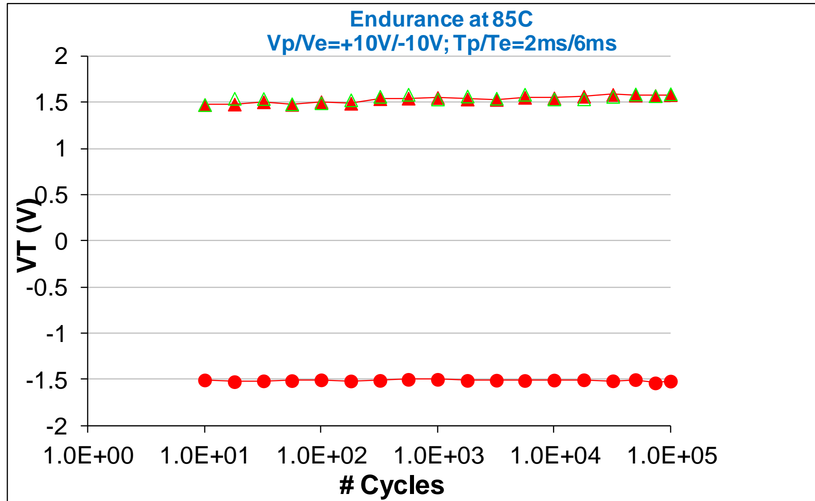

Figure 4. Endurance Characteristics of Cypress SONOS FET

The endurance characteristics show that Vt shift is less than 100 mV after 100k program/erase cycles at 85°C.

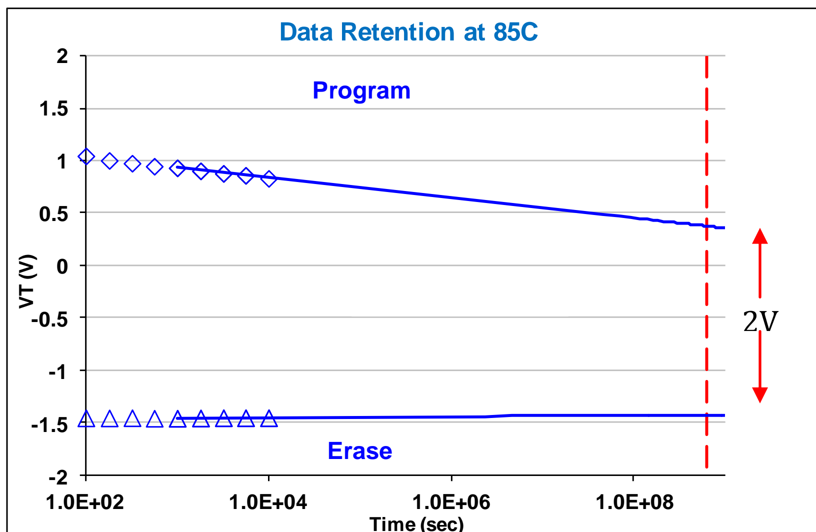

Figure 5. Retention Characteristics of Cypress SONOS FET

The data retention characteristics show that the Cypress SONOS cell has an EOL window of 2 V. The requirement for a successful read on Cypress SONOS flash macros is 0.4 V, so there is plenty of margin with our typical EOL window of 2 V.

SONOS Integration

The SONOS module can easily be embedded into a logic process flow or an SRAM process flow with the addition of three masking layers. The integration scheme includes a dual-gate oxide process, which enables the chip to be compatible with multiple supply voltages. SONOS has been integrated into a state-of-the-art logic process flow, which includes salicided junctions, stress layers, and Cu-low K interconnects. The simplicity of the SONOS structure using the same poly gate as the logic and simply replacing the gate oxide with an ONO allows easy scaling even to High-K metal-gate processes with either the gate first or gate last approaches. Low thermal budget of SONOS integration ensures a negligible impact on the electrical parameters of existing CMOS devices. This has been demonstrated on devices, design IP and a high density SRAM product in the 65nm technology. This means that all design IP of the original CMOS platform can be used with SONOS integration.

The integration of SONOS into a logic process flow uses significantly fewer masks than FG flash, as shown in Table 2:

Process Step	Embedded Floating-Gate NVM	CY Embedded SONOS NVM
Wells	+3 Masks	
Floating gate patterning	+1 Mask	
Other NVM related patterning	+2 Masks	+1 Mask
High voltage oxidation/patterning	+1 Mask	+1 Mask
NVM /source/drain implants	+2 Masks	+1 Mask
High voltage source/drain implant	+2 Masks	
Total masking step adder	+11 Masks	+3 Masks

Table 2. Comparison of Additional Masks for FG Flash and SONOS Embedded into Logic Flow

In this integration, the high-voltage circuits of the flash macro such as charge pumps and latches are designed using Drain Extended MOSFETs (DENMOS and DEPMOS) which use the I/O gate oxide but are capable of handling high voltages of up to 9.0 V between Drain and Source. In 65-nm SONOS, these DEMOS FETs are designed by using the existing implants without the need for additional DEMOS specific implant masks. However, in advanced nodes such as 55 nm / 40 nm / 28 nm, the baseline implants are not adequate for the DEMOS FETs to meet the specifications. Therefore, dedicated implants are added for the DEMOS FETs (one each for DENMOS and DEPMOS). This will need the addition of two more masks, which brings the total additional masks to 5.

SONOS Technology in High-Volume Manufacturing

The 0.35-µm SONOS Technology (known as S4) and 130-µm SONOS technology (known as S8) have been running in high volume in multiple fabs. The main products made in Cypress Minnesota fab are Programmable System on Chip (PSoC) and Nonvolatile SRAM (NVSRAM). In this fab, we have manufactured nearly 150,000 S4 wafers and 200,000 S8 wafers. The SONOS technology has been transferred to Grace Semiconductor Manufacturing Company (GSMC), China. In GSMC, 125,000 S4 wafers and 150,000 S8 wafers have been fabricated for Cypress PSoC products. Hua Hong NEC, another foundry in China, licensed the S8 technology and has manufactured nearly 500,000 wafers since 2007 to make Smart Card products.

SONOS Integration into Foundry 65-nm Process

SONOS was integrated into the 65-nm CMOS baseline process in the foundry without impacting the baseline process. Device matching data before and after SONOS integration is shown in Figure 6 and Figure 7.

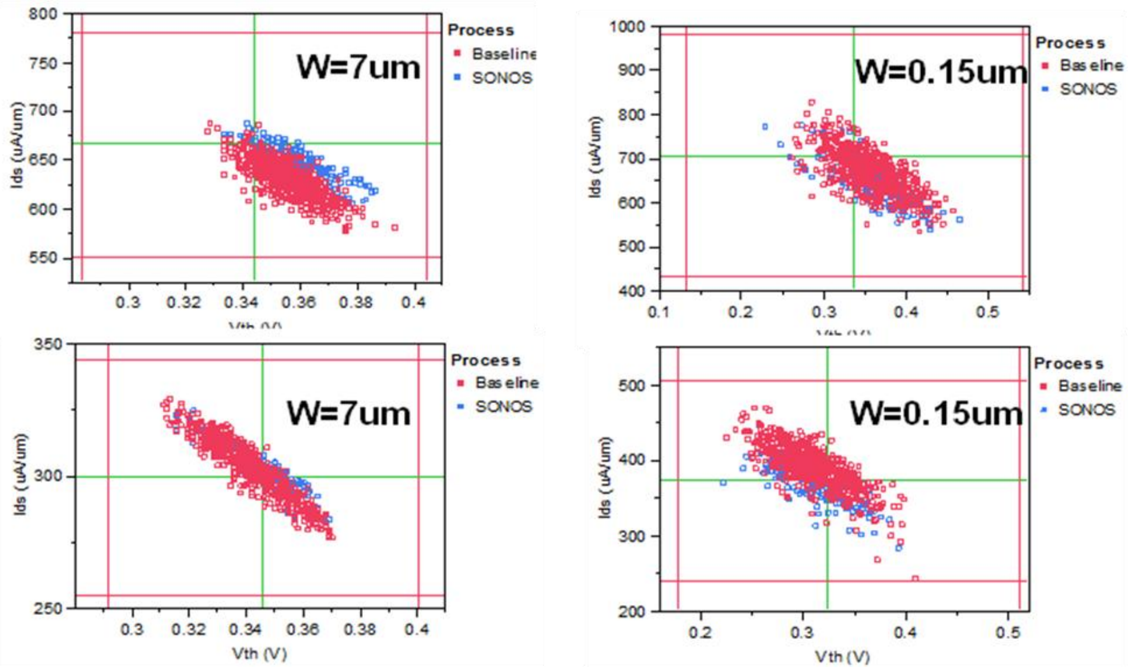

Figure 6. Matching of CMOS FETs in SONOS Technology to the CMOS FETs in Logic Baseline

To determine the impact of SONOS module on yield, the SONOS process (deposition, etch, and thermal budget) was added to a 72-Mbit SRAM product and the yields were comparable. The data below shows that there is no negative impact of the SONOS process steps.

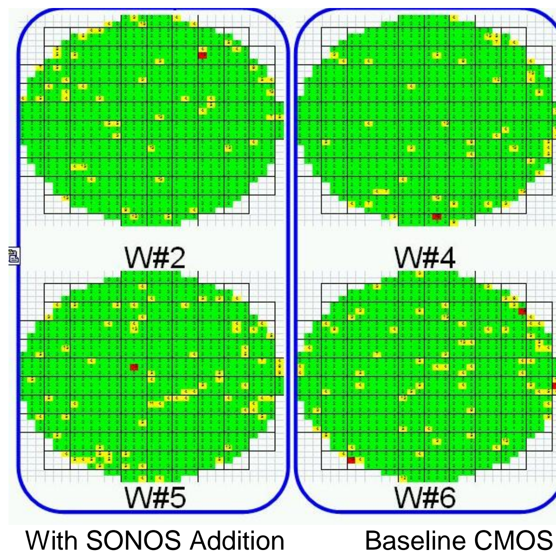

Figure 7. 72-Mbit SRAM Product Yield Comparison

65-nm SONOS Cell Operation

The 65-nm SONOS cell is also a 2T cell and is programmed and erased in the same way as the 130-nm cell (using FN tunneling). This SONOS technology has the additional challenge of lower program/erase voltages. Since the baseline 65-nm technology has an I/O voltage of 2.5 V, the program/erase voltages are limited to about 8 V. The SONOS stack has been re-optimized to meet the performance and reliability requirements at these program/erase voltages, which is the lowest reported for this node. Typical bias voltages for the 7.5-V program/erase operations are shown in Table 3.

	CG	BL	SL	P-Well
Erase	-3.5V	4.0V	4.0V	4.0V
Program	4.0V	-3.5V	-3.5V	-3.5V
Read	0V	0.6V	0V	0V

Table 3. Voltages for 65-nm SONOS Operation

Cypress 65-nm SONOS Performance Specifications

The optimized Cypress 65-nm SONOS technology has the following performance specifications:

Program time: 1-5 ms; Erase time: 2.5-10 ms

Program Vt (BOL): 1.2 V; Erase Vt (BOL): –1 V

Minimum selected erased cell read current: 3 μA

Maximum selected programmed cell read current: 100 nA/cell

Maximum unselected cell leakage at 100°C: 3.9 nA/cell

Data retention: 10 years at 85°C

Endurance: 100k cycles

65-nm SONOS Cell Reliability

The 65-nm SONOS cell has been characterized for reliability in the same way as for the 130-nm technology. The measured endurance and retention characteristics are shown in Figure 8 and Figure 9:

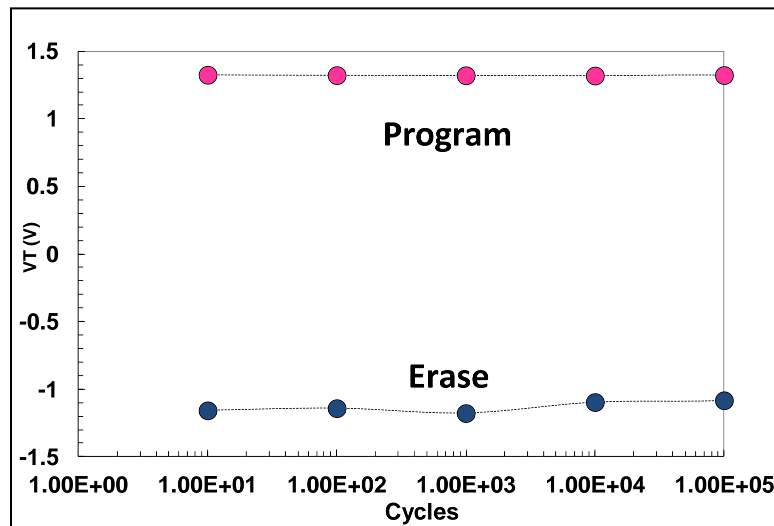

Figure 8. SONOS Cell Endurance @ 85° C

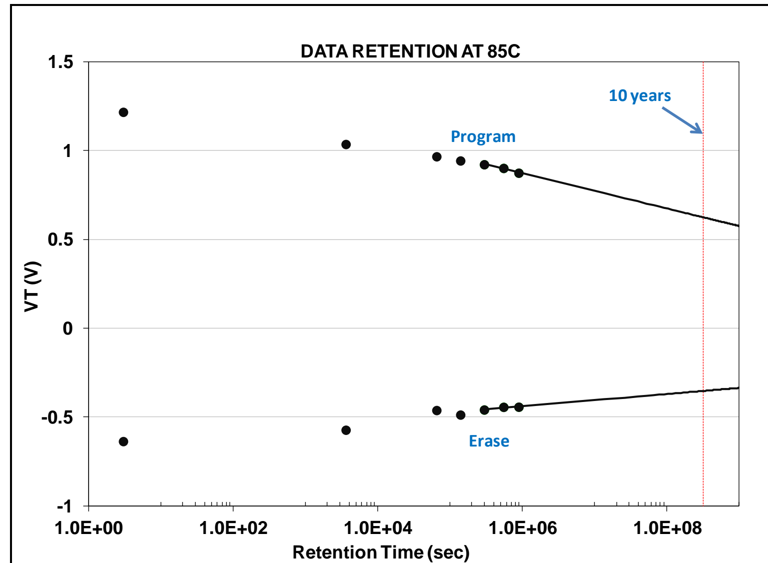

Figure 9. SONOS Cell Retention @ 85° C

65-nm SONOS Cell-Based Flash Macro

A 4.5-Mbit memory macro based on the 65-nm SONOS cell has been designed, fabricated and qualified. This macro has consistently yielded more than 90% after data retention bake. The key performance parameters of this macro are given in Table 4:

Cell Technology	65 nm – 2T SONOS (FN/FN)
Supply Voltage	1.08 V to 1.32 V
Ambient Temperature	-40°C to 85°C
Read Cycle Time (1.08 V)	28 ns
Endurance	100k cycles
Data Retention	10 years
Program/Erase Current	10 mA (for 4k bits)
NVM Mask Adders	3 (use 2.5 V IO GOX for HV)

Table 4. Flash Macro Key Performance Parameters

This flash macro has also been characterized for endurance and retention. The measured reliability data is shown in Figure 10 and Figure 11.

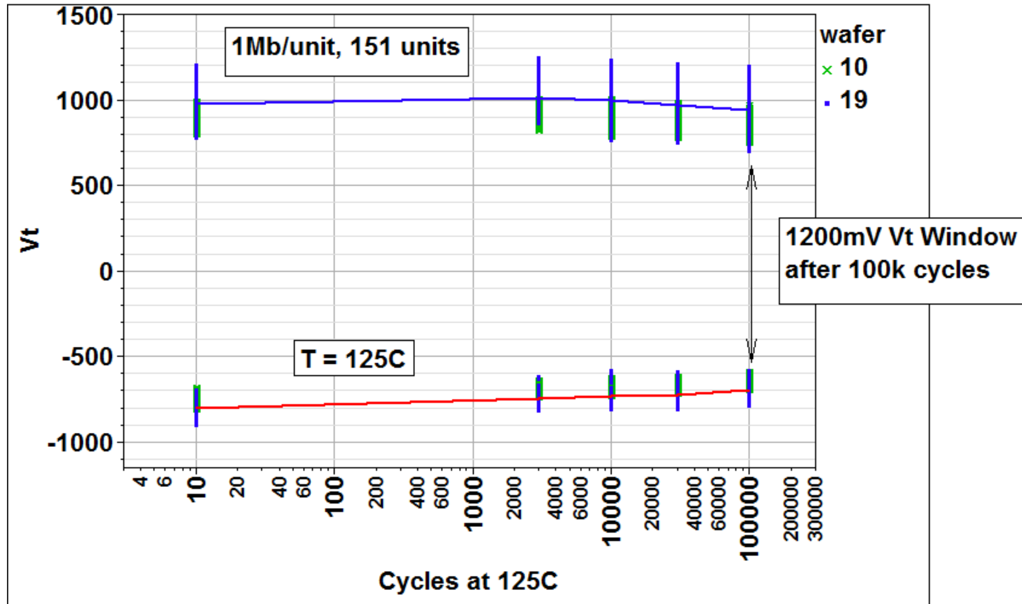

Figure 10. Endurance Performance of Flash Macro

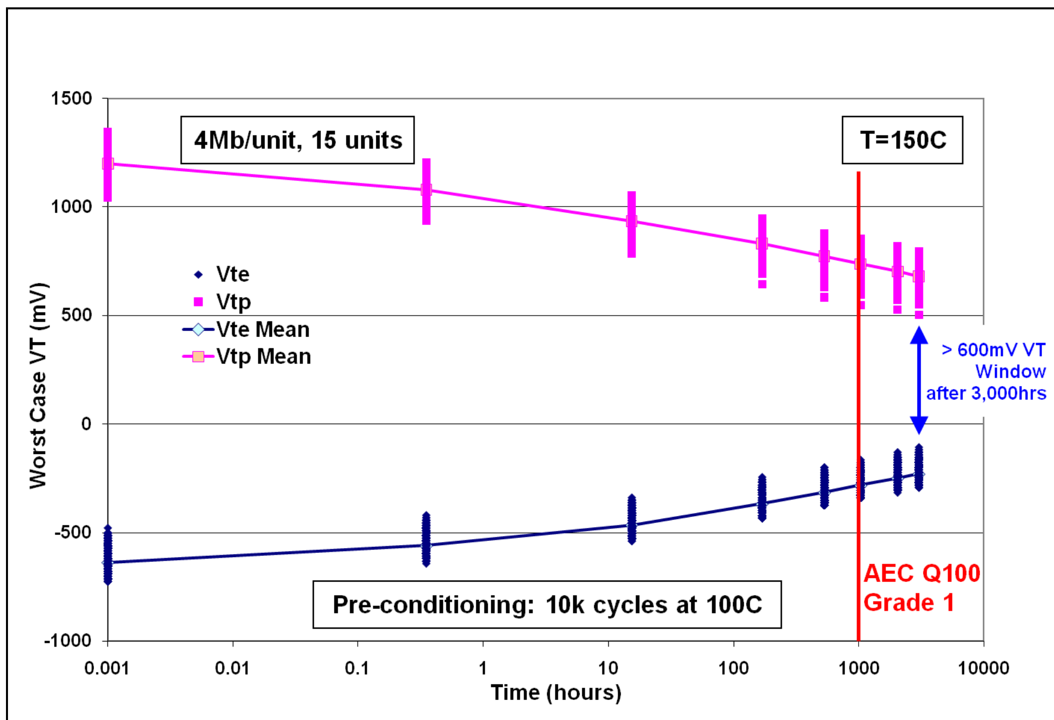

Figure 11. Data Retention Performance of Flash Macro

Scaling of SONOS Technology

SONOS technology is scalable to advanced nodes such as 55 nm, 40 nm, and 28 nm since there is no additional complexity introduced by the SONOS devices for lithography. SONOS cells at the 28 nm node have been successfully fabricated using a 28 nm Poly-SiON baseline process at a foundry. Initial characterization shows acceptable Vts and program/erase window (Figure 12).

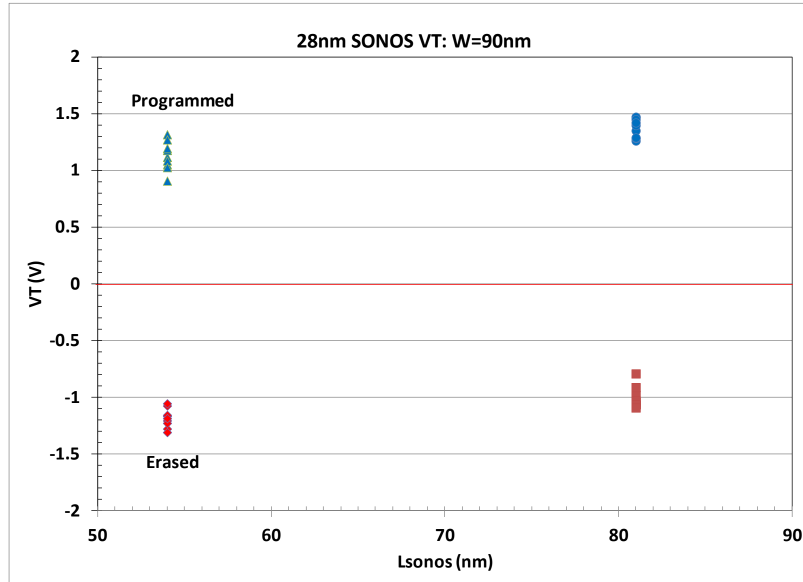

Figure 12. Threshold Voltages of 28-nm SONOS FETs

Summary

The Cypress SONOS technology is a highly reliable, low-cost solution to embed nonvolatile memory into logic/mixed signal platforms. The CMOS design IP is unaffected by SONOS integration. This technology has excellent endurance and data retention and is scalable to more advanced technology nodes. There are no known barriers to embedding Cypress's SONOS technology into High-K metal-gate processes.

Cypress Semiconductor
198 Champion Court
San Jose, CA 95134-1709
Phone: 408-943-2600
Fax: 408-943-4730
http://www.cypress.com

© Cypress Semiconductor Corporation, 2008-2016. The information contained herein is subject to change without notice. Cypress Semiconductor Corporation assumes no responsibility for the use of any circuitry other than circuitry embodied in a Cypress product. Nor does it convey or imply any license under patent or other rights. Cypress products are not warranted nor intended to be used for medical, life support, life saving, critical control or safety applications, unless pursuant to an express written agreement with Cypress. Furthermore, Cypress does not authorize its products for use as critical components in life-support systems where a malfunction or failure may reasonably be expected to result in significant injury to the user. The inclusion of Cypress products in life-support systems application implies that the manufacturer assumes all risk of such use and in doing so indemnifies Cypress against all charges.

PSoC Designer™ and Programmable System-on-Chip™ are trademarks and PSoC® is a registered trademark of Cypress Semiconductor Corp. All other trademarks or registered trademarks referenced herein are property of the respective corporations.

This Source Code (software and/or firmware) is owned by Cypress Semiconductor Corporation (Cypress) and is protected by and subject to worldwide patent protection (United States and foreign), United States copyright laws and international treaty provisions. Cypress hereby grants to licensee a personal, non-exclusive, non-transferable license to copy, use, modify, create derivative works of, and compile the Cypress Source Code and derivative works for the sole purpose of creating custom software and or firmware in support of licensee product to be used only in conjunction with a Cypress integrated circuit as specified in the applicable agreement. Any reproduction, modification, translation, compilation, or representation of this Source Code except as specified above is prohibited without the express written permission of Cypress.

Disclaimer: CYPRESS MAKES NO WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, WITH REGARD TO THIS MATERIAL, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. Cypress reserves the right to make changes without further notice to the materials described herein. Cypress does not assume any liability arising out of the application or use of any product or circuit described herein. Cypress does not authorize its products for use as critical components in life-support systems where a malfunction or failure may reasonably be expected to result in significant injury to the user. The inclusion of Cypress' product in a life-support systems application implies that the manufacturer assumes all risk of such use and in doing so indemnifies Cypress against all charges.

Use may be limited by and subject to the applicable Cypress software license agreement.