Infineon SONOS non-volatile memory technology

Abstract

This white paper discusses the key features of SONOS-based non-volatile memory (NVM) technology, its performance, and reliability. It also highlights the suitability of SONOS-based NVM for Rad-Hard applications.

This document provides an overview of the Infineon SONOS (Silicon-Oxide-Nitride-Oxide-Silicon) technology (originally developed by Cypress, which was acquired by Infineon in 2020), its features, and its integration into CMOS process flow for embedded non-volatile memory (NVM) applications.

Table of contents

1 Introduction

The demand for embedded flash memory has grown steeply over the years as many new applications emerged in consumer electronics (touchscreens, smart cards, bank cards, mobile payment, e-passport, etc.) in addition to the industrial system-on-chip (SOC) designs. The memory content of these applications has been increasing steadily due to greater requirements on system performance and code/data storage. Initially, embedded floating-gate flash memory was widely used despite the large number of mask adder required (which translates to manufacturing cost) and the challenges in preserving the baseline CMOS device models (which affects compatibility with existing IPs).

Since then, SONOS emerged as a compelling alternative for embedded flash memory at a significantly lower cost. SONOS has been known as an NVM technology since the 1980s. In the early days, however, it was not very successful in competing against the floating-gate technology due to higher programming voltage and less competitive data retention at high temperatures. Infineon solved these obstacles through charge-trap engineering in the SONOS non-volatile memory stack and demonstrated 10-year retention at up to 125°C (ambient temperature) with robust margins. Today, Infineon SONOS technology is used in many products not only from Infineon but also numerous other companies (through technology licensing). These products include smart cards, standalone NOR Flash, FPGAs, and microcontrollers. SONOS is also robust for Radiation Hardness, making it ideal for Rad-Hard products. Recently, a new and exciting application has opened for SONOS: Analog NVM for neuromorphic computing in artificial intelligence (AI) edge applications. Infineon is working to optimize the operation of the SONOS technology to tackle this exciting field.

2 Flash memories

Non-volatile memories retain stored information even if the power supply to the memory is switched off. There are different types of non-volatile 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. In non-volatile memories, information is stored in a specific layer of the memory device as charge or resistance states.

2.1 SONOS memory technology

The SONOS memory uses an insulating film such as silicon nitride with traps as the charge-storage layer. The traps in the nitride capture the charge carriers injected from a semiconductor channel and retain the charge. This charge determines the Vt/I_{DS} of the MOSFET in which the charge-storage nitride is part of the gate dielectric. 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 defects and more robust for data retention. The key advantage of the current SONOS technology is the significantly lower voltages required for program/erase operations compared to floating-gate flash. In Infineon SONOS, by controlling the deposition parameters in the ONO formation, the ONO stack is engineered to maximize the charge-trapping efficiency during erase and program operations and minimize the charge loss during retention. With such a stack, Infineon SONOS meets automotive-grade reliability specs.

3 Overview of Infineon SONOS technology

Infineon SONOS is an embedded NVM technology that integrates a highly reliable SONOS transistor into a CMOS process flow with very few added mask layers and minimal impact on the electrical parameters of the existing CMOS devices. Infineon SONOS has been in volume production since 2001 with over two billion chips shipped to date at nodes ranging from 350 nm to 40 nm. 22 nm SONOS technology with High K – Metal Gate (HKMG) is currently in development.

3.1 Infineon SONOS transistor and cell

The heart of the Infineon SONOS technology is the SONOS transistor shown in Figure 1. This is a MOS transistor with an ONO stack as the gate dielectric. The ONO is designed to provide the required program and erase speeds and excellent reliability. The trap distribution in the film is engineered to provide outstanding endurance and retention characteristics. Furthermore, the stack is suitable for high-volume manufacturing with excellent process control. The SONOS transistor shares most of the key process steps with CMOS transistors. Hence, many regions of the SONOS transistor including 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 the floating gate.

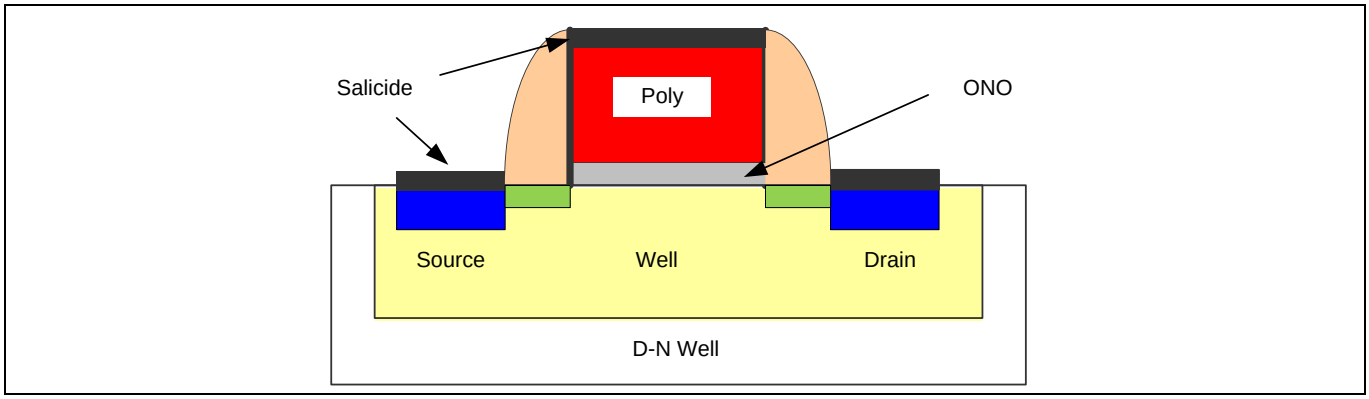

Figure 1 SONOS transistor cross-section

Figure 1 is a SONOS transistor that is fabricated using a typical logic CMOS process flow. The Infineon SONOS technology currently offers multiple cell options to fit into different applications, trading off cell (and macro) size verses access time. Similarly, on the most advanced technology nodes that use HKMG, the SONOS transistor exploits all of the key benefits of HKMG.

3.1.1 SONOS cell operation

Infineon SONOS is based on two-transistor (2T) cell comprised of a SONOS transistor in series with an NMOS select transistor. The cell is programmed by applying a positive field between the gate and the substrate to inject electrons from the substrate into the charge-storage layer by Fowler-Northeim (FN) tunneling. Erase operation is performed by applying a negative field, which causes FN tunneling of holes from the substrate to the charge-storage layer. The required program and erase bias levels between the gate and the substrate are obtained by applying appropriate voltage polarities to the gate and the p-well. The program speed is 1-3 ms, and the erase speed is 2.5 ms-10 ms, depending on cell option and macro architecture.

To achieve very aggressive flash macro size, Infineon has introduced a new SONOS array architecture based on “Common Source Line,” which has one common source line for an entire row of cells instead of the conventional “Dedicated Source Line” that has one source line for each column of cells. This approach is implemented at the 55 nm node and below.

3.1.2 SONOS integration

The first SONOS technology developed was on 0.35 um CMOS baseline process flow and qualified in 2001. The next node was 130 nm, qualified in 2008; this node continues to generate high volume of products even today. Beyond 130 nm, the SONOS integration moved to foundry partners. At the 40 nm node, SONOS was integrated with a foundry's 40 nm CMOS process flow with only five additional masks. The integration scheme includes a dual-gate oxide (which enables compatibility with multiple supply voltages), mobility-enhancing layers, and low-K dielectric films in metal-interconnect stack. The simplicity of the SONOS structure using the same poly gate as the CMOS devices and simply replacing the gate oxide with the ONO stack allows easy scaling even to HKMG processes with either gate-first or gate-last approaches. Low thermal budget of SONOS integration also ensures negligible impact on the electrical parameters of existing CMOS devices. Consequently, all design IPs of the original CMOS platform can be re-used with SONOS integration. Figure 2 shows TEM of a 40 nm SONOS cell.

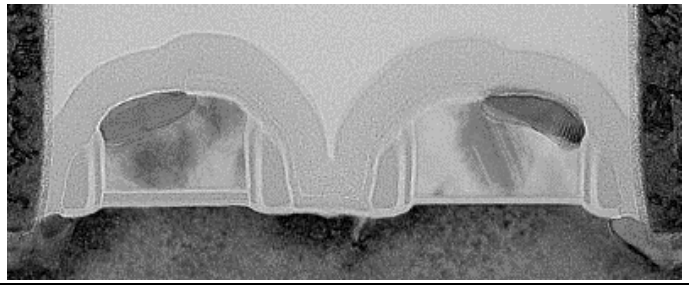

Figure 2 40 nm SONOS cell cross-section

3.1.2.1 40 nm SONOS cell operation

As previously described, the 40 nm SONOS technology uses a Common Source Line cell which is significantly smaller. The cell is programmed and erased in the same way as in the preceding nodes (using FN tunneling). The SONOS stack has been optimized to meet the performance and reliability requirements while generating only up to +/-4V to create the net field needed for program and erase. The program and erase characteristics of the 40 nm cell are shown in Figure 3.

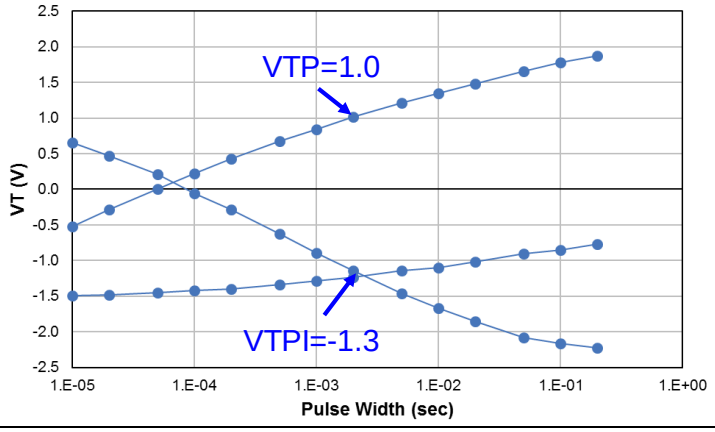

Figure 3 Program and erase characteristics of Cypress

The Cypress 40 nm SONOS performance specifications are almost the same as those of the 65 nm technology.

3.1.2.2 40 nm SONOS cell-based flash macro

Key performance parameters of 40 nm SONOS macro are listed in Table 1, and Vt distributions after cycling and retention bake are shown in Figure 4 and Figure 5.

Table 1 Flash macro key performance parameters

Parameters	Value
CMOS process	40 µLP
Density	1 Mb to 16 Mb
Output data width	32-, 64-, or 128-bit
Dual power supply	0.8 V-1.21 V and 1.62 V-3.63 V
Operating temperature (T_j)	-40°C to 125°C
Read access time	25 ns (1.10 ± 0.11 V)
Read current	61 mA/MHz w/ 32-bit data-out
Write endurance	100k cycles
Data retention	10 years at 100°C

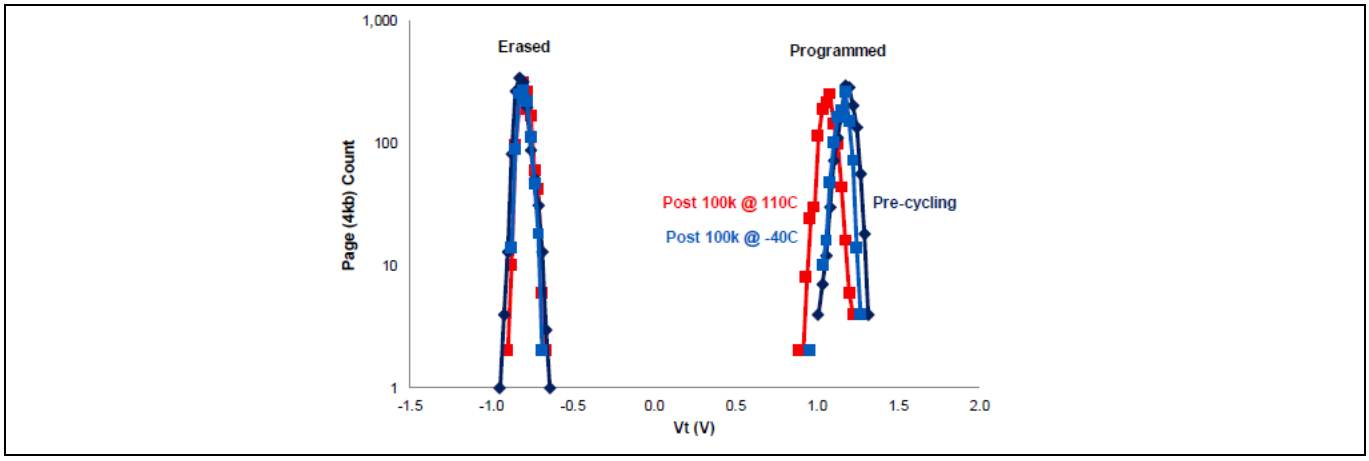

Figure 4 Endurance performance of 40 nm SONOS flash macro

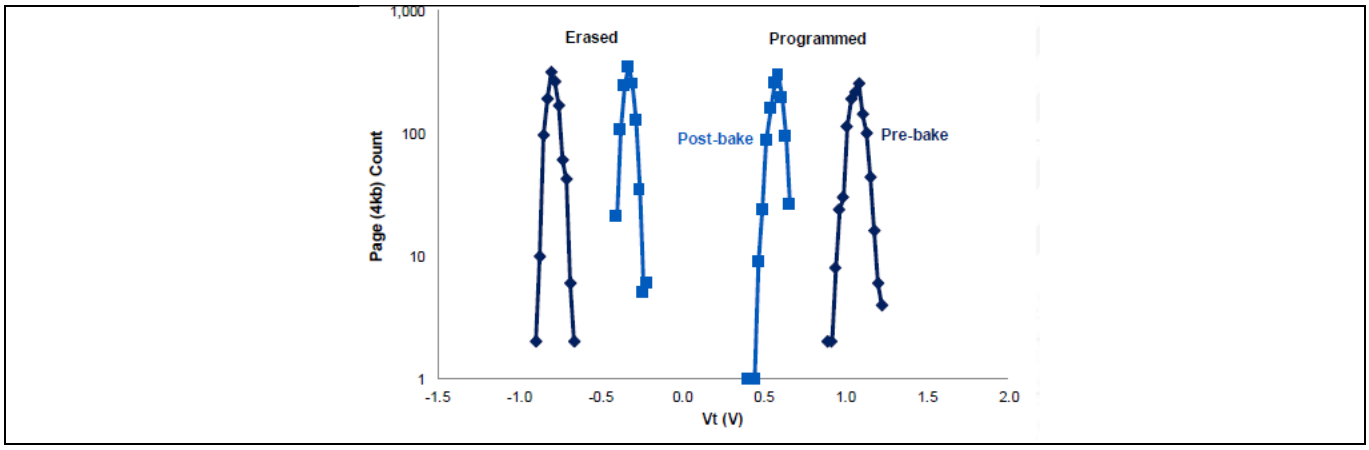

Figure 5 Data retention performance of flash macro

3.2 SONOS memory for Rad-Hard applications

The 40 nm SONOS memory technology has been used to develop standalone NOR memory for aerospace applications which demand radiation hardness, and a 512 Mb NOR memory product is fully qualified. It demonstrated the required radiation-hardness capability, exceeding the total ionizing dose (TID) and single event latch-up (SEL) requirements.

4 Scaling of SONOS technology – 28 nm and 22 nm

The SONOS technology is scalable to advanced nodes such as 28/22 nm, as there is no additional complexity introduced by the SONOS devices for any of the process steps.

28 nm Poly–SiON

SONOS cells at the 28 nm node have been successfully fabricated using a foundry 28 nm Poly-SiON process flow. This technology is running in high-volume production to manufacture FPGAs of very large memory size. More than ten products are currently being manufactured.

22 nm HKMG

Integration of SONOS into 22 nm HKMG baseline process is currently ongoing. TEM cross section of a SONOS cell in the 22 nm gate-last HKMG technology is shown in Figure 6. Vt and program/erase window are preserved relative to the 40 nm node (Figure 7).

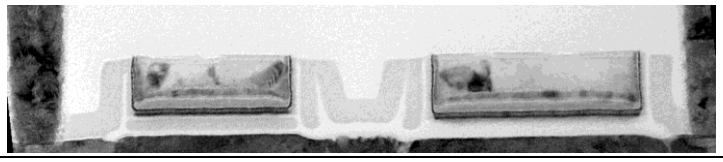

Figure 6 22 nm SONOS cell cross-section

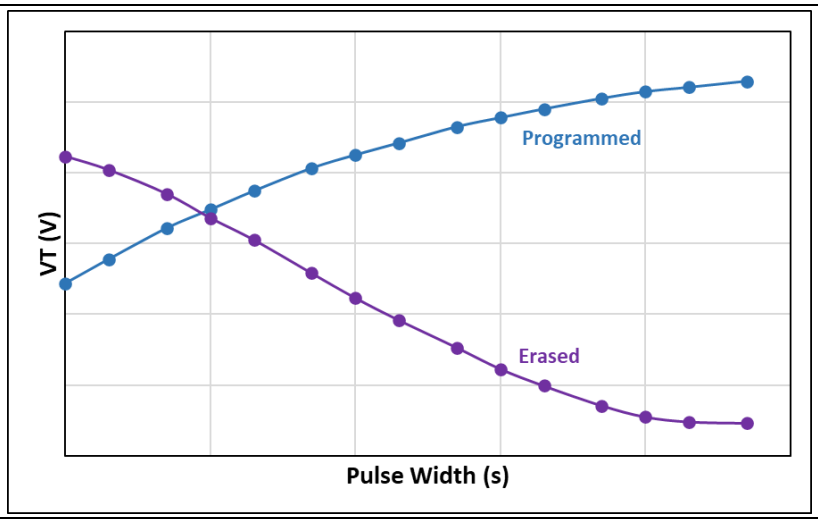

Figure 7 Program and erase characteristics of 22 nm HKMG SONOS

5 Summary

The Infineon SONOS technology is a highly reliable, low-cost solution to embed nonvolatile memory into logic or mixed-signal platforms. The CMOS design IPs are unaffected by the SONOS integration. This technology has excellent endurance and data retention and is scalable to advanced technology nodes including HKMG processes. SONOS has also been demonstrated to be highly suitable for Analog NVM in neuromorphic computing applications; this new and exciting application will undoubtedly lead to even broader adoption of SONOS in the coming years.

Revision history

Document revision	Date	Description of changes
*F	2024-08-30	Template update

Trademarks
All referenced product or service names and trademarks are the property of their respective owners.

Edition 2024-08-30
Published by

Infineon Technologies AG
81726 Munich, Germany

Do you have a question about this document?
Email: erratum@infineon.com

Document reference
001-46554 Rev. *F

Important notice
This document is for information purposes only and any information given herein shall in no event be regarded as a warranty, guarantee or description of any functionality, conditions and/or quality of our products or any suitability for a particular purpose with regard to the technical specifications of our products, we kindly ask you to refer to the relevant product data sheets provided by us. Our customers and their technical departments are required to evaluate the suitability of our products for the intended application.

We reserve the right to change this document and/or the information given herein at any time.

Warnings
Due to technical requirements, our products may contain dangerous substances. For information on the types in question please contact your nearest Infineon Technologies office.
Except as otherwise explicitly approved by us in a written document signed by authorized representatives of Infineon Technologies, our products may not be used in any life endangering applications, including but not limited to medical, nuclear, military, life critical or any other applications where a failure of the product or any consequences of the use thereof can result in personal injury.